SUPPLEMENTAL METHODS

Participants

We obtained GWAS results in the form of summary statistics (p-values and z-scores). Data on schizophrenia (SCZ) were retrieved from Psychiatric Genomics Consortium (PGC) through collaboration (1). The SCZ dataset consisted of 53 386 individuals with schizophrenia or schizoaffective disorder and 77 258 controls of European descent (1). Further, we used GWAS results on CVD phenotypes, including the CVD risk factors BMI (n=795 640) (2), waist-to-hip ratio (WHR) adjusted for BMI (n=694 649) (3), triglycerides (TG, n=1 320 016) (4), total cholesterol (TC) (n=1 320 016) (4), low-density lipoprotein (LDL) cholesterol (n=1 320 016) (4), high-density lipoprotein (HDL) cholesterol (n=1 320 016) (4), systolic and diastolic blood pressure (n=745 820-757 601) (5), type 2 diabetes mellitus (T2D, n=231 426) (6), smoking initiation (a binary variable indicating whether an individual had ever smoked regularly, n=1 232 091) (7), cigarettes per day (CigPerDay, n=337 334) (7), and coronary artery disease (CAD, n=332 477, including 71 602 CAD cases and 260 875 controls) (8). We restricted analyses to samples of European ancestry to limit population stratification. There was no overlap between participants in the CVD GWASs and the SCZ sample by inspecting the substudies that the GWAS data were retrieved from. We did not have access to individual genotype data and were therefore not able to identify potential overlapping individuals across studies. However, based on the sample description, it is highly unlikely that there is a significant number of overlapping individuals. Details about the inclusion criteria, genotyping and phenotype characteristics, see the original publications (1, 2, 4-8).

MiXeR

We applied causal mixture models (9, 10) to the GWAS summary statistics, using the MiXeR tool (https://github.com/precimed/mixer). For each SNP, i, univariate MiXeR models its

additive genetic effect of allele substitution, β_i, as a point-normal mixture, $\beta_i = (1-\pi_1)N(0,0) + \pi_1 N(0,\sigma_\beta^2)$, where π_1 represents the proportion of non-null SNPs (`polygenicity`) and σ_β^2 represents variance of effect sizes of non-null SNPs (`discoverability`). Then, for each SNP, j, MiXeR incorporates LD information and allele frequencies for M=9,997,231 SNPs extracted from 1000 Genomes Phase3 data by LD score regression software (10, 11), and estimate the expected probability distribution of the signed test statistic, $z_j = \delta_j + \epsilon_j = N\sum_i \sqrt{H_i} r_{ij}\beta_i + \epsilon_j$, where N is sample size, H_i indicates heterozygosity of i-th SNP, r_{ij} indicates allelic correlation between i-th and j-th SNPs, and $\epsilon_j \sim N(0,\sigma_0^2)$ is the residual variance. Further, the three parameters, $\pi_1, \sigma_\beta^2, \sigma_0^2$, are fitted by direct maximization of the likelihood function. The number of trait-influencing variants (i.e. variants with pure genetic effects not induced by LD) is estimated as $M\pi_1$, where M=9,997,231 gives the number of SNPs in the reference panel.

In the cross-trait analysis, MiXeR models additive genetic effects as a mixture of four components, representing null SNPs in both traits (π_0); SNPs with a specific effect on the first and on the second trait (π_1 and π_2, respectively); and SNPs with non-zero effect on both traits (π_{12}). In the last component, MiXeR models variance-covariance matrix as $\mathbf{\Sigma_{12}} = \begin{bmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$ where ρ_{12} indicates correlation of effect sizes within the shared component, and σ_1^2 and σ_2^2 correspond to the discoverability parameter estimated in the univariate analysis of the two traits. After fitting parameters of the model, genetic correlation is calculated as $r_g = \frac{\rho_{12}\pi_{12}}{\sqrt{(\pi_1+\pi_{12})(\pi_2+\pi_{12})}}$. Further information is available in (9). After estimating the size of the shared and unique components, the dice coefficient (DC) was calculated to represent the overall extent of genetic overlap using the formula DC = $\frac{2\pi_{12}}{\pi_1+\pi_2+2\pi_{12}}$.

To evaluate model fit, i.e. the ability of the MiXeR to accurately predict the actual GWAS data, we constructed modelled vs. actual conditional Q-Q plots (Figures S3-11). The conditional Q-Q plots show observed versus expected $-\log_{10}$ p-values in the primary trait as a function of the significance of association with a secondary trait at the level of $p \leq 0.1$, $p \leq 0.01$, and $p \leq 0.001$, with successive leftward deflection of SNP strata with higher significance in the secondary (conditional) trait indicating cross-trait enrichment. Model fit is demonstrated in these conditional Q-Q plots if the data Q-Q plots (solid lines) are closely reproduced by the model predictions (dashed lines) across all p-value strata (9). Interestingly, scenarios without polygenic overlap may show an enrichment because GWAS p-values depend on allele frequency and LD structure, although this effect is smaller than enrichment arising due to shared variants (9). Furthermore, model fit was also assessed using negative log-likelihood plots (9), which visualize the performance of the best model versus models with minimum and maximum polygenic overlap (Figures S3-11). More specifically, the best model with polygenic overlap estimated with MiXeR was compared with two models – a model with least possible overlap and a model with maximum possible overlap. In the negative log-likelihood plot (Figures S3-11), the minimum model is represented by the point furthest to the left, the maximum model is represented by the point furthest to the right, and the lowest point on the curve (y-axis) indicates better model fit (9). Support for the MiXeR model is a clearly defined minimum on the negative log-likelihood curve.

To filter situations with insufficiently powered GWAS summary statistics, we use Akaike information criterion ($AIC = 2k - 2\ln L$), where k is the number of free parameters in the model, L is the value of the likelihood function, and n is the effective number of SNPs used in optimization procedure. We calculate the difference between *AIC* for the full bivariate model, $k = 3$, and AIC for the reduced bivariate model, $k = 2$, due to π_{12} being constrained to smallest or largest possible ($\pi_{12}^{min} = r_g\sqrt{\pi_1^u\,\pi_2^u}$ and $\pi_{12}^{max} = \min(\pi_1^u, \pi_2^u)$, respectively).

A positive value of AIC indicates that GWAS summary statistics have enough information to distinguish the custom polygenic overlap, as shown on the MiXeR Venn diagrams, from the constrained models with minimal (π_{12}^{min}) and maximum (π_{12}^{max}) polygenic overlap.

Conditional False Discovery Rate

The 'enrichment' seen in the conditional Q-Q plots can be directly interpreted in terms of true discovery rate (TDR = 1 – false discovery rate (FDR)) (12). More specifically, for a given p-value cutoff, the FDR is defined as

$$FDR(p) = \pi_0 F_0(p) / F(p), \qquad [1]$$

where π_0 is the proportion of null SNPs, F_0 is the null cumulative distribution function (cdf), and F is the cdf of all SNPs, both null and non-null (13). Under the null hypothesis, F_0 is the cdf of the uniform distribution on the unit interval [0,1], so that Eq. [1] reduces to

$$FDR(p) = \pi_0 p / F(p), \qquad [2]$$

The cdf F can be estimated by the empirical cdf $q = N_p / N$, where N_p is the number of SNPs with p-values $\leq$ p, and N is the total number of SNPs. Replacing F by q in Eq. [2], we get

$$\text{Estimated } FDR(p) = \pi_0 p / q, \qquad [3]$$

which is biased upwards as an estimate of the FDR (14). Replacing π_0 in Equation [3] with unity gives an estimated FDR that is further biased upward;

$$q^* = p / q, \qquad [4]$$

If π_0 is close to one, which is probably true for most GWASs, the increase in bias from Eq. [3] is minimal. Therefore, the quantity 1 – p/q, is biased downward and thus a conservative estimate of the TDR. Referring to the Q-Q plots, we see that q* is equivalent to the nominal p-value divided by the empirical quantile, as defined previously. We can thus read the FDR estimate directly off the Q-Q plot as

$$-\log_{10}(q^*) = \log_{10}(q) - \log_{10}(p), \qquad [5]$$

demonstrating that the estimated FDR is directly related to the horizontal shift of the curves in the Q-Q plots from the expected line x = y, i.e. a larger shift corresponds to a smaller FDR.

Conditional Q-Q plots

Q-Q plots compare a nominal probability distribution against an empirical distribution. In the presence of all null relationships, nominal p-values form a straight line on a Q-Q plot when plotted against the empirical distribution. For SCZ and CVD phenotype SNPs and for each categorical subset (strata), $-\log_{10}$ nominal p-values were plotted against $-\log_{10}$ empirical p-values (conditional Q-Q plots). Leftward deflections of the observed distribution from the projected null line illustrate increased tail probabilities in the distribution of test statistics (z-scores) and consequently an over-abundance of low p-values compared to that expected by chance, also called 'enrichment'. This is illustrated in Figures S12-13.

Under large-scale testing paradigms, such as GWAS, we can calculate quantitative estimates of likely true associations from the distributions of summary statistics (13, 15). Conditional Q-Q plots of nominal p-values from GWAS summary statistics visualizes this enrichment of statistical association relative to that expected under the global null hypothesis. The usual Q-Q curve has the nominal p value, denoted by "p", as the y-ordinate and the corresponding value of the empirical cdf, denoted by "q", as the x-ordinate. Under the global null hypothesis the theoretical distribution is uniform on the interval [0,1]. As is common in GWAS, we instead plot $-\log_{10}$ p against $-\log_{10}$ q to emphasize tail probabilities of the theoretical and empirical distributions. Therefore, genetic enrichment is illustrated with a leftward shift in the Q-Q curve, corresponding to a larger fraction of SNPs with nominal $-\log_{10}$ p-value greater than or equal to a given threshold. Conditional Q-Q plots are constructed by creating subsets of SNPs based on levels of an auxiliary measure for each SNP, and computing Q-Q plots separately for each level. If SNP enrichment is captured by variation in

the auxiliary measure, this is expressed as successive leftward deflections in a conditional Q-Q plot as levels of the auxiliary measure increase. We constructed conditional Q-Q plots of empirical quantiles of nominal -$\log_{10}$ values for SNP association for all SNPs, and for subsets (strata) of SNPs determined by the nominal p-values of their association with the conditional phenotypes, and vice versa. In particular, we computed the empirical cumulative distribution (cdf) of nominal p-values for a given phenotype for all SNPs and for SNPs with significance levels below the indicated cut-offs for the conditional phenotypes ($-\log_{10}(p) \geq 1$, $-\log_{10}(p) \geq 2$, $-\log_{10}(p) \geq 3$ corresponding to $p < 0.1$, $p < 0.01$, $p < 0.001$ respectively). The nominal p-values ($-\log_{10}(p)$) are plotted on the y-axis, and the empirical quantiles ($-\log_{10}(q)$, where q=1-cdf(p)) are plotted on the x-axis. To assess for polygenic effects below the standard GWAS significance threshold, we focused the conditional Q-Q plots on SNPs with nominal $-\log_{10}(p) < 7.3$ (corresponding to $p > 5x10^{-8}$). We controlled for spurious enrichment by calculating all conditional Q-Q plots after random pruning averaged over 500 iterations. At each iteration, one SNP in every LD block (defined by an $r^2 > 0.1$) was randomly selected and the empirical cdfs were computed using the corresponding p-values.

Detection of SNPs using conditional and conjunctional FDR

The FDR can be interpreted as the probability that a SNP is null given that its p-value is as small as or smaller than its observed p-value. The conditional FDR (condFDR) is an extension of the standard FDR, which incorporates information from GWAS summary statistics of a second phenotype to adjust its significance level. The condFDR is defined as the probability that a SNP is null in the first phenotype given that the p-values in the first and second phenotypes are as small as or smaller than the observed ones. It is important to note that ranking SNPs by the standard FDR or by p-values gives the same ordering of SNPs. In contrast, ranking SNPs by condFDR will reorder SNPs when the primary and secondary

phenotypes are genetically related. The conjunctional FDR (conjFDR) is defined as the posterior probability that a SNP is null for either phenotype or both simultaneously, given that its p-values for association with both phenotypes are as small as or smaller than the observed p-values (16-20). A conservative estimate of the conjFDR is obtained by the maximum condFDR for a given SNP after repeating the condFDR procedure for both traits and inverting their roles (21). Given that complex correlations in regions with intricate LD can bias FDR estimation (22), we excluded SNPs in the extended major histocompatibility complex (MHC) and chromosome 8p23.1 (genome build 19 locations chr6: 25119106–33854733 and chr8: 7242715–12483982, respectively) and SNPs in LD (r^2>0.1) with such SNPs before fitting the FDR models. Accordingly, we avoided artificially inflated genetic enrichment owing to the LD structure of these regions (23). This is the standard procedure in cond/conjFDR studies of SCZ and related disorders that show strong associations with these genomic regions (23-25). The procedure involves excluding MHC and 8p23.1 before creating the conditional Q-Q plots, which allows us to observe cross-trait enrichment that is not driven by the strong associations within these regions. However, SNPs within these regions are not removed from the cond/conjFDR analyses per se because SNPs in these regions are given a cond/conjFDR value based on their p-value and may have high biological relevance. Furthermore, MiXeR provides accurate estimates in the presence of realistic LD structure. Accordingly, MHC was also excluded when performing MiXeR analyses as the intricate LD structure creates difficulties in providing reliable model estimates (9). The 8p23.1 region is characterized by strongly correlated SNPs, yet less complex LD pattern, and thus MiXeR is able to model this LD pattern (9). P-values were corrected for inflation using a genomic inflation control procedure (16).

Genomic loci definition

We defined independent genomic loci using the FUMA, an online tool for functional mapping of genetic variants (http://fuma.ctglab.nl/) (27). Summary statistics from the GWASs on SCZ and CVD phenotypes were used as input for FUMA. First, *independent significant SNPs* were identified as SNPs with condFDR < 0.01 and independent from each other at LD $r^2 < 0.6$. Secondly, *lead SNPs* were identified by retaining those independent significant SNPs that were independent from each other at $r^2 < 0.1$. Next, *distinct genomic loci* were identified by merging physically overlapping lead SNPs (LD blocks < 250 kb apart) selecting a SNP with the most significant p-value as a lead SNP of the merged locus. Borders of the genomic loci were determined by identifying all SNPs in LD ($r^2 \geqq 0.6$) with one of the independent significant SNPs in the locus. The region containing all of these *candidate SNPs* was regarded as a single independent genomic locus. All LD information was calculated from the 1000 Genomes Project reference panel (28).

Effect sizes and genetic correlation

Effect size (z-scores) of the shared SNPs were obtained from the original summary statistics (see original publications (1-8)). We estimated the genetic correlation using LD score regression (29). LD score regression was estimated using the Python-based package available at https://github.com/bulik/ldsc. The procedure is described in here: https://github.com/bulik/ldsc/wiki/Heritability-and-Genetic-Correlation.

Identification of novel SCZ loci

We identified novel SCZ loci by comparing the identified loci at conjFDR <0.05 with the loci reported in the most recent SCZ GWAS (1), other SCZ GWASs and cond/conjFDR analyses, and the NHGRI-EBI catalog (30).

Validation: Consistency of genetic effects in independent samples

The probability of replicating individual loci at genome-wide significance is low due to weak genetic effects. Therefore, we determined the number of lead SNPs in the identified shared loci (conjFDR<0.05) whose effect directions were the same in replication samples (i.e. sign concordance), following a similar procedure as described previously (31, 32). There was sign concordance if the lead SNP had concordant effect directions in the replication samples. If lead SNPs were missing in the independent datasets for conjFDR analyses, we replaced them with the next most significant candidate SNP which was present in the independent datasets. We employed a one-sided exact binomial test of significance under the null hypothesis that significant concordance was randomly distributed (31, 32). In addition, we identified lead SNPs (or next most significant candidate SNPs) which were nominally significant at $P < 0.05$ in each independent sample (31, 32).

Functional annotation

We used FUMA (27), an online annotation platform (http://fuma.ctglab.nl/) to functionally annotated all candidate SNPs in the genomic loci with a condFDR or conjFDR value<0.10 having an $r^2 \geqq 0.6$ with one of the independent significant SNPs. SNPs were annotated with Combined Annotation Dependent Depletion (CADD) scores (33), RegulomeDB (34) scores, and chromatin states (35, 36). The CADD score is a deleterious score of variants computed by integrating 63 functional annotations (33). The higher the score, the more deleterious. A CADD score above 12.37 is the threshold to be potentially pathogenic (33). The RegulomeDB score is a categorical score to guide interpretation of regulatory variants (34). It is based on information from eQTLs and chromatin marks, ranging from 1a to 7 with lower scores indicating a higher likelihood of having a regulatory function. Scores are as follows: 1a=eQTL + Transcription Factor (TF) binding + matched TF motif + matched DNase

Footprint + DNase peak; 1b=eQTL + TF binding + any motif + DNase Footprint + DNase peak; 1c=eQTL + TF binding + matched TF motif + DNase peak; 1d=eQTL + TF binding + any motif + DNase peak; 1e=eQTL + TF binding + matched TF motif; 1f=eQTL + TF binding / DNase peak; 2a=TF binding + matched TF motif + matched DNase Footprint + DNase peak; 2b=TF binding + any motif + DNase Footprint + DNase peak; 2c=TF binding + matched TF motif + DNase peak; 3a=TF binding + any motif + DNase peak; 3b=TF binding + matched TF motif; 4=TF binding + DNase peak; 5=TF binding or DNase peak; 6=other; 7=Not available (34).

The chromatin state represents the accessibility of genomic regions (every 200bp) with 15 categorical states predicted by a hidden Markov model based on 5 chromatin marks for 127 epigenomes in the Roadmap Epigenomics Project (35). A lower state indicates increased accessibility, with states 1-7 referring to open chromatin states. We annotated the minimum chromatin state across tissues to SNPs. The 15-core chromatin states as suggested by Roadmap are as follows: 1=Active Transcription Start Site (TSS); 2=Flanking Active TSS; 3=Transcription at gene 5' and 3'; 4=Strong transcription; 5= Weak Transcription; 6=Genic enhancers; 7=Enhancers; 8=Zinc finger genes & repeats; 9=Heterochromatic; 10=Bivalent/Poised TSS; 11=Flanking Bivalent/Poised TSS/Enh; 12=Bivalent Enhancer; 13=Repressed PolyComb; 14=Weak Repressed PolyComb; 15=Quiescent/Low. Standardized SNP effect sizes were calculated for the most impactful SNPs by transforming the sample size-weighted meta-analysis *Z* score, in line with Zhu et al. (36).

Furthermore, using FUMA (27), we linked candidate SNPs to genes applying either of three gene mapping strategies: 1) positional mapping to align SNPs to genes based on their physical proximity (i.e., within a 10kb window), 2) expression quantitative trait locus (eQTL) mapping to match cis-eQTL SNPs to genes whose expression is associated with allelic variation at the SNP level, and 3) chromatin interaction mapping to link SNPs to genes based

on three-dimensional DNA–DNA interactions between each SNP's genomic region and nearby or distant genes. The eQTL analyses were corrected for multiple comparisons using an FDR threshold of 0.05. FUMA contains Hi-C data of over 21 tissue/cell types including human brain tissue (https://fuma.ctglab.nl/tutorial#chromatin-interactions). We used an FDR of 1 x 10^{-6} to define significant chromatin interactions based on the suggestion by Schmitt et al.(37). After excluding mapped genes in complex LD regions (MHC and 8p23.1), we performed Gene Ontology (GO) gene-set analysis and pathway analysis using FUMA (38). In addition, the Molecular Signatures Database evaluated enrichment in immunological signature gene sets (39). Analyses were corrected for multiple comparisons (Bonferroni correction).

The gene-mapping approach involving three strategies has high sensitivity, but increases false positives that bias gene-set and pathway analyses. To reduce false positives, we performed mapping of SNPs to genes based on *physical proximity* as studies suggest that the nearest gene is often the causal gene (40). Next, we performed gene-set and pathway analysis of these positionally mapped genes. Since we identified a large number of loci, we focused on *lead* SNPs in this gene-mapping analysis, further prioritizing specificity over sensitivity to ensure more robust gene-set analyses. We compared the results with findings from gene-set and pathway analysis of genes mapped to candidate SNPs using all three strategies.

Cross-CVD trait analyses

We assessed the genetic overlap between the CVD phenotypes. In particular, we estimated the genetic correlation (LDSR) between each CVD phenotype and performed bivariate MiXeR analyses of most CVD phenotypes. We chose CVD traits for MiXeR based on the type of risk factor, selecting one representative of each category of CVD risk factor (e.g. blood lipids,

blood pressure, obesity). The selection of CVD risk factors for MiXeR was also influenced by the polygenicity and heritability of the traits, which affect the model's ability to provide accurate estimates.

Genomic SEM

We performed Genomic structural equation modelling (Genomic SEM) (41, 42) to investigate the influence of CVD risk factors on the relationship between SCZ and CAD. We followed the analysis procedure in (41). We applied a multiple regression model to estimate the effect of SCZ on CAD and vice versa controlling for a third variable (e.g. BMI) (41). In addition, Genomic SEM was used for mediation analysis to distinguish between direct and indirect effects on the relationship between SCZ and CAD (41, 42). Indirect effects indicate the influence of the CVD risk factors on this relationship. For further information, see the original Genomic SEM publication (42) and method descriptions by Wormington et al. (41).

SUPPLEMENTAL RESULTS

MiXeR results

Univariate MiXeR estimated different SNP-based heritability of SCZ and the various CVD phenotypes. SCZ (h^2_{SNP}=0.38) and BMI (h^2_{SNP}=0.22) possessed highest h^2_{SNP}, followed by DBP, SB, T2D, WHR and lipids (h^2_{SNP}=0.11-0.16) (Table S100). Smoking initiation, CigPerDay and CAD had the lowest SNP-based heritability (h^2_{SNP}=0.04-0.07).

Univariate Q-Q plots demonstrate that MiXeR-based predictions provide accurate estimates of the data plots (Figure S1). In addition, the Q-Q plots with SNPs partitioned according to MAF & LD score, as expected, show a stronger signal for SNPs with higher MAF and higher LD score for all traits (Figure S2). The model predictions for SCZ and BMI follow the same pattern, indicating that the model correctly captures the dependency of GWAS association statistics on MAF and LD (Figure S2). For other traits, we noted minor departure between

predicted and observed GWAS curves, mostly prominent for SNPs in the bin with low MAF and low LD score, which tend to have higher z-score than what is predicted by the MiXeR model. This indicates enrichment in additive genetic effects towards low-MAF and low-LD SNPs, in line with previously published results (43-46). However, we observe that with current MiXeR model the discrepancy between observed and predicted z-score distribution in low-MAF and low-LD bin is only prominent towards the tail of z-score distribution, while the majority of z-scores were still described accurately by the MiXeR model. Also, the positive AIC and BIC values for all univariate analyses indicate significant improvement of the causal mixture model over an infinitesimal model (pi=1) (Table S100), indicating that the model captures unique pattern of polygenicity for all traits considered in the analysis. The difference between observed and predicted z-score distributions towards the tail of z-score distribution is relevant for fine-mapping and functional annotation applications, and incorporating these aspects of genetic architecture is subject of current development of MiXeR 2.0 model.

MiXeR results, including number of shared and unique trait-influencing variants and corresponding standard error, are presented in Figure 1. Table S100 presents the proportion of SCZ-linked variants influencing CVD, and vice versa. The variants explain 90% of the SNP heritability in each phenotype. Using MiXeR, we discovered extensive polygenic overlap between SCZ and smoking initiation, sharing 8.6K (SE=0.5K) out of 12.5K variants involved, as illustrated by the Venn diagram (Figure 1a). The shared variants represent 89.6% of the genetic variants influencing SCZ (9.6K) and 77.5% of the variants underlying smoking initiation (11.1K) (Table S100). MiXeR also revealed polygenetic overlap between SCZ and BMI, sharing 8.1K variants (SE=0.5K) out of 12.6K variants, as visualized in the Venn diagram (Figure 1b). The shared variants with BMI represent 83.5% of the genetic variants influencing SCZ, and 73.6% of variants influencing BMI (11.0K) (Table S100). Further, MiXeR identified genetic overlap with SBP, sharing 2.4K (SE=0.2K) out of 11.7K variants

(Figure 1c). The shared variants with SBP represent 24.7% of the variants influencing SCZ and 54.5% of the variants influencing SBP (4.4K) (Table S100). MiXeR estimated 1.7K (SE=0.2K) shared variants with DBP out of 11.8K variants (Figure 1d). The shared variants with DBP make up 17.7% of the variants influencing SCZ and 43.6% of the variants influencing DBP (3.9K) (Table S100). Further, MiXeR estimated 1.6K (SE=0.3K) shared variants between SCZ and T2D out of 10.3K variants (Figure 1e), constituting 16.7% of SCZ influencing variants and 69.6% of variants underlying T2D (2.3K) (Table S100). In addition, MiXeR estimated that 1.4K (SE=0.2K) out of 10.1K variants are shared between SCZ and HDL (Figure 1f), 1.2K (SE=0.1K) out of 9.9K variants are shared with TG, 0.3K (SE=0.1K) out of 10.2K variants are shared with LDL, and 0.3K (SE=0.1K) out of 9.9K variants are shared variants with TC (Figure S3). The overlapping variants with HDL constitute 14.4% of SCZ-linked variants and 77.8% of variants influencing HDL (Table S100). The overlapping variants with TG constitute 12.4% of SCZ-linked variants and 85.7% of variants influencing TG (Table S100). The shared variants with TC and LDL represent 3.1% of SCZ-linked variants and 33.3% of variants associated with TC and LDL (Table S100). MiXeR predicted that 0.5K (SE=0.2K) out of 10.5K variants are jointly influencing SCZ and CAD (Figure S4). The overlapping variants represent 5.2% of SCZ influencing variants, while 34.5% of variants influencing CAD (Table S100). The model estimated that 1.2K (SE=0.3K) out of 11.2K variants are influencing both SCZ and WHR (Figure S10). The shared variants represent 12.1% of SCZ influencing variants and 43.5% of WHR influencing variants (Table S100).

The bivariate MiXeR estimates adequately model the GWAS data, as indicated by the model-based Q-Q plots following the actual Q-Q plots (Figures S3-10). The negative log-likelihood plots also illustrated adequate model fit, as indicated by a clearly defined minimum on the curve (Figures S3-10). Further, AIC demonstrated sufficient power of the model (Table S101). The positive AIC values indicate that the MiXeR model is adequately powered to

differentiate the estimated polygenic overlap from minimum possible overlap (best vs. min. overlap) and maximum possible overlap (best vs. max. overlap) (Table S101). MiXeR was not applied for CigPerDay due to inadequate model fit, as shown in the negative log-likelihood plots not displaying a clear minimum on the curve (Figure S11). MiXeR does not seem to reliably estimate the number of variants influencing CigPerDay (Figure S11).

We estimated DC, i.e. the percentage of shared variants between SCZ and each CVD phenotype out of the total number of SNPs influencing both traits. The DC (mean, SD) for SCZ and smoking was 83.0% (4.0), SCZ and BMI was 78% (4.9), SCZ and SBP was 33.8% (3.4), SCZ and DBP was 25.1% (3.3), SCZ and T2D was 27.1% (4.0), SCZ and HDL was 23.9% (3.2), SCZ and TG was 20.9% (2.6), SCZ and LDL was 6.0% (2.3), SCZ and TC was 5.9% (2.3), and SCZ and CAD was 9.9% (3.4), and SCZ and WHR was 19.0% (4.7).

Visualizing cross-trait enrichment

In the conditional Q-Q plots, we observed enrichment in SCZ SNPs as a function of the significance of associations with CVD phenotypes (Figure S12), indicating polygenic overlap. The reverse conditional Q-Q plots also illustrated enrichment in CVD phenotypes given associations with SCZ (Figure S13). This indicates polygenic overlap between SCZ and CVD phenotypes.

Conditional FDR results

To increase statistical power, we leveraged the pleiotropic enrichment using condFDR analysis and re-ranked SCZ SNPs conditional on their association with CVD phenotypes, and vice versa. At condFDR<0.01, we identified 362 loci associated with SCZ conditional on smoking initiation; 325 loci conditional on SBP; 317 loci conditional on DBP, 303 conditional on WHR, 307 conditional on CigPerDay, 332 loci conditional on TG, 331 loci

conditional on HDL, 279 loci conditional on LDL, 272 loci conditional on TC, 299 loci conditional on BMI, 291 loci conditional on T2D, and 246 loci conditional on CAD (Tables S1-12). Next, we identified multiple loci associated with CVD phenotypes conditional on associations with SCZ (Tables S13-24).

Effect directions of shared lead SNPs between SCZ and CVD phenotypes

We evaluated the directionality of allelic effects of the shared lead SNPs between the phenotypes by investigating their z-scores (Tables S25-36 and S101). We discovered the same effect direction in 105/304 loci (34.5%) in SCZ and BMI, 202/293 (68.9%) in SCZ and smoking initiation, 81/129 (62.8%) in SCZ and CigPerDay, 140/294 (47.6%) in SCZ and SBP, 124/ 259 (47.9%) in SCZ and DBP, 69/147 (46.9%) in T2D and SCZ, 145/307 (47.2%) in TG and SCZ, 169/304 (55.6%) in HDL and SCZ, 81/158 (51.3%) in LDL and SCZ, 95/176 (53.5%) in TC and SCZ, 20/35 (57.1%) in CAD and SCZ, and 101/193 (52.3%) in WHR and SCZ (Tables S25-36).

Validation: Concordance and significance of shared loci in independent samples

The results from replication analysis are provided in Tables S41-47. Some of the lead SNP or nearby candidate SNPs lacked summary statistics in the replication dataset (Not available, NA) and thus their replicability could not be assessed. 223 of 287 lead SNPs in loci shared between SCZ and BMI (77%; exact binomial $P = 5.8 \times 10^{-28}$) and 205 of 271 lead SNPs in loci shared between SCZ and SBP (76%; exact binomial $P = 4.71 \times 10^{-18}$) were sign concordant in the independent SCZ GWAS (Tables S41-42). Of these SNPs, 73 and 58, respectively, had $P < 0.05$ in the replication SCZ sample (Tables S41-42). 204 of 268 lead SNPs in loci shared between SCZ and smoking initiation (76%; exact binomial $P = 1.66 \times 10^{-18}$) and 93 of 121 lead SNPs in loci shared between SCZ and CigPerDay (77%; exact

binomial $P = 1.22 \times 10^{-9}$) were sign concordant in the independent SCZ GWAS (Tables S43-44). Of these SNPs, 44 and 26, respectively, had $P < 0.05$ in the replication SCZ sample (Tables S43-44). 226 of 278 lead SNPs in loci shared between SCZ and TG (81%; exact binomial $P = 2.47 \times 10^{-27}$) and 212 of 275 lead SNPs in loci shared between SCZ and HDL (77%; exact binomial $P = 2.51 \times 10^{-20}$) were sign concordant in the independent SCZ GWAS (Tables S45-46). Of these SNPs, 57 and 50, respectively, had $P < 0.05$ in the replication SCZ sample (Tables S45-46). 27 of 34 lead SNPs in loci shared between SCZ and CAD (79%; exact binomial $P = 0.0004$) were sign concordant in the independent SCZ GWAS (Table S47). Of these SNPs, nine had $P < 0.05$ in the replication SCZ sample (Table S47).

233 of 288 lead SNPs in loci shared between SCZ and BMI (81%; exact binomial $P = 5.98 \times 10^{-20}$) and 268 of 294 lead SNPs in loci shared between SCZ and SBP (91%; exact binomial $P = 4.15 \times 10^{-52}$) were sign concordant in the independent GWASs for BMI and SBP (Tables S41-42). Of these SNPs, 74 and 91, respectively, had $P < 0.05$ in these replication samples for BMI and SBP (Tables S41-42). 186 of 242 lead SNPs in loci shared between SCZ and smoking initiation (77%; exact binomial $P = 8.75 \times 10^{-18}$) and 75 of 113 lead SNPs in loci shared between SCZ and CigPerDay (66%; exact binomial $P = 0.0003$) were sign concordant in the replication samples for smoking phenotypes (Tables S43-44). Of these SNPs, 40 and 16, respectively, had $P < 0.05$ in independent GWAS samples for smoking (Tables S43-44). 205 of 294 lead SNPs in loci shared between SCZ and TG (70%; exact binomial $P = 5.39 \times 10^{-12}$) and 199 of 289 lead SNPs in loci shared between SCZ and HDL (69%; exact binomial $P = 1.03 \times 10^{-10}$) were sign concordant in the replication samples for lipids (Tables S45-46). Of these SNPs, 41 and 41, respectively, had $P < 0.05$ in the replication samples for lipids (Tables S45-46). 24 of 29 lead SNPs in loci shared between SCZ and CAD (83%; exact binomial $P = 0.0002$) were sign concordant in the replication CAD sample (Table S47). Of these SNPs, three had $P < 0.05$ in the replication sample for CAD (Table S47).

Note that the replication GWAS samples were considerably smaller than the discovery samples, which reduced the power to detect significant lead SNPs. The replication samples were around 1/2 (SCZ, SBP) to 1/9 (smoking, lipids, CAD) of the primary sample sizes. Given the smaller sample sizes, the number of replicated lead SNPs with nominal significance is as expected. The consistency of associations in replication datasets is similar to that of other GWASs of complex traits (31, 32).

Gene-mapping of candidate SNPs

Gene-mapping of candidate SNPs: By applying FUMA, we mapped candidate SNPs jointly associated with SCZ and CVD phenotypes to protein-coding genes. We linked the candidate SNPs in the shared loci between SCZ and BMI to 2047 protein-coding genes (Table S60). Positional gene-mapping aligned SNPs to 799 genes, eQTL gene-mapping implicated 1081 genes, and chromatin interaction mapping indicated 1394 genes (Table S60). 357 genes were implicated by all three mapping strategies, providing higher credibility for these genes (Table S60). We linked the candidate SNPs in the shared loci between SCZ and WHR to 2246 protein-coding genes (Table S61). Positional gene-mapping aligned SNPs to 847 genes, eQTL gene-mapping implicated 1065 genes, and chromatin interaction mapping indicated 1786 genes (Table S61). 438 genes were implicated by all three mapping strategies, providing higher credibility for these genes (Table S61). Furthermore, we linked the candidate SNPs in the shared loci between SCZ and smoking initiation to 1895 protein-coding genes, of which 357 genes were implicate by all three strategies (Table S62). Positional gene-mapping aligned SNPs to 797 genes, eQTL gene-mapping implicated 999 genes, and chromatin interaction mapping indicated 1328 genes (Table S62). We linked the candidate SNPs in the shared loci between SCZ and CigPerDay to 899 protein-coding genes, of which 161 genes were implicate by all three strategies (Table S63). Positional gene-mapping aligned SNPs 364 genes, eQTL

gene-mapping implicated 450 genes, and chromatin interaction mapping indicated 627 genes (Table S63). We linked the candidate SNPs in the shared loci between SCZ and TG to 2212 protein-coding genes, of which 336 genes were implicate by all three strategies (Table S64). Positional gene-mapping aligned SNPs to 874 genes, eQTL gene-mapping implicated 1165 genes, and chromatin interaction mapping indicated 1499 genes (Table S64). We linked the candidate SNPs in the shared loci between SCZ and HDL to 2158 protein-coding genes, of which 313 genes were implicate by all three strategies (Table S65). Positional gene-mapping aligned SNPs to 774 genes, eQTL gene-mapping implicated 1145 genes, and chromatin interaction mapping indicated 1416 genes (Table S65). We linked the candidate SNPs in the shared loci between SCZ and LDL to 1511 protein-coding genes, of which 189 genes were implicate by all three strategies (Table S66). Positional gene-mapping aligned SNPs to 466 genes, eQTL gene-mapping implicated 805 genes, and chromatin interaction mapping indicated 963 genes (Table S66). We linked the candidate SNPs in the shared loci between SCZ and TC to 1741 protein-coding genes, of which 243 genes were implicate by all three strategies (Table S67). Positional gene-mapping aligned SNPs to 573 genes, eQTL gene-mapping implicated 895 genes, and chromatin interaction mapping indicated 1206 genes (Table S67). The candidate SNPs in the shared loci between SCZ and SBP were mapped to 2558 protein-coding genes, of which 555 genes were implicate by all three strategies (Table S68). Positional gene-mapping aligned SNPs to 1080 genes, eQTL gene-mapping implicated 1359 genes, and chromatin interaction mapping indicated 1046 genes (Table S68). We linked the candidate SNPs in the shared loci between SCZ and DBP to 2460 protein-coding genes, of which 595 genes were implicate by all three strategies (Table S69). Positional gene-mapping aligned SNPs 1121 genes, eQTL gene-mapping implicated 1365 genes, and chromatin interaction mapping indicated 1788 genes (Table S69). We linked the candidate SNPs in the shared loci between SCZ and T2D to 1445 protein-coding genes, of which 269 genes were

implicate by all three strategies (Table S70). Positional gene-mapping aligned SNPs to 517 genes, eQTL gene-mapping implicated 780 genes, and chromatin interaction mapping indicated 1053 genes (Table S70). We linked the candidate SNPs in the shared loci between SCZ and CAD to 451 protein-coding genes, of which 75 genes were implicate by all three strategies (Table S71). Positional gene-mapping aligned SNPs to 147 genes, eQTL gene-mapping implicated 264 genes, and chromatin interaction mapping indicated 318 genes (Table S71). The gene-mapping of the candidate SNPs jointly associated with SCZ and CVD phenotypes largely implicated genes expressed in brain tissue (adult and fetal cortex) and immune cells (Tables S60-71).

Gene-set and pathway analysis of genes nearest to lead SNPs

We performed gene-set and pathway analyses of genes nearest to lead SNPs. The analyses implicated that genes nearest to lead SNPs jointly associated with SCZ and BMI were enriched in several Gene Ontology (GO) terms, including "regulation of transmembrane transport", "regulation of cell development", "regulation of synaptic plasticity" and several neuronal gene-sets (Table S72). No significant GO-terms were enriched among the genes nearest to lead SNPs shared between SCZ and WHR, but a pathway termed "KEGG Alzheimer's Disease" appeared significant (Table S73). Gene-set analysis of the genes nearest to the lead SNPs shared between SCZ and smoking initiation also indicated several GO terms, with a predominance of gene-sets related to neurodevelopment, including "central nervous system development" and "neuron development" (Table S74). The GO gene-set analysis of the nearest mapped genes for SCZ and lipids implicated "positive regulation of RNA biosynthetic process" and "cell-cell signaling" and several neuronal (e.g. "neuron differentiation"), synaptic (e.g. "synapse organization") and nucleic acid binding (e.g. "sequence specific DNA binding") gene-sets (Tables S75-77). The GO gene-sets enriched

with genes nearest to lead SNPs associated with SCZ and SBP/DBP included "homophilic cell adhesion via plasma membrane adhesion molecules" and "cell-cell adhesion via plasma membrane adhesion molecules" as well as several neuronal and synaptic gene-sets (Tables S78-79). In addition, immunological gene-sets were enriched with the genes nearest to lead SNPs associated with SCZ and TG, HDL, SBP (Tables S75-76; 78). There were no significantly enriched gene-sets with genes nearest to lead SNPs shared between SCZ and T2D and CAD.

The analyses implicated pathways overrepresented among the genes nearest to lead SNPs shared between SCZ and smoking initiation, including "Pathways Affected in Adenoid Cystic Carcinoma" (Table S74). Pathway analysis of the genes mapped to SCZ and TG indicated "MAPK Signaling Pathway" and "Brain-Derived Neurotrophic Factor (BDNF) signaling pathway" (Table S75). Pathway analysis of the genes mapped to SCZ and SBP and DBP indicated "MAPK Signaling Pathway", "Brain-Derived Neurotrophic Factor (BDNF) signaling pathway". "Energy Metabolism" and "Adherens junction" (Tables S78-79). There were no pathways significantly overrepresented with genes nearest to lead SNPs shared between SCZ and BMI, TC, LDL, HDL, T2D and CAD.

Genes-set analysis of genes mapped to candidate SNPs

Gene-set analysis revealed that mapped genes for SCZ and BMI were enriched in four Gene Ontology (GO) terms, including "hemophilic cell adhesion via plasma membrane adhesion molecules" and "MRNA binding" (Table S80). The mapped genes for SCZ and WHR were also enriched in several GO terms, including "homophilic cell adhesion via plasma membrane adhesion molecules" and "amide biosynthetic process" (Table S81). The mapped genes for SCZ and smoking initiation were enriched in several GO terms, with a predominance of gene-sets related to neurodevelopment, including "central nervous system development" (Table S82). The genes mapped to SCZ and lipids were enriched in GO terms associated with inter-

and intracellular processes and structures, including “homophilic cell adhesion via plasma membrane adhesion molecules”, “intracellular transport” and “mitochondrion” (Tables S83-86). Mapped genes for SCZ and SBP and DBP were most significantly enriched in the GO terms “homophilic cell adhesion via plasma membrane adhesion molecules”, “synapse organization” and several neuronal gene-sets, including “neuron projection” (Tables S87-88). Analysis of mapped genes for SCZ and T2D implicated enriched gene-sets involving molecular functions, including “adenyl nucleotide binding” and “ribonucleotide binding” (Table S89). There were no significant GO results for the genes mapped to SCZ and CigPerDay and CAD (Table S90). However, the analyses showed enrichment in several immunological signature gene-sets across all CVD phenotypes (Tables S80-91).

Pathway analysis of genes mapped to candidate SNPs

Pathway analysis of the genes mapped to SCZ and CigPerDay indicated “glycine, serine and threonine metabolism” (Table S92). Pathway analysis of the genes mapped to SCZ and T2D indicated “ppara pathway” (Table S93). This pathway involves gene regulation by Peroxisome Proliferators via Peroxisome proliferator-activated receptor-α (PPARα). Analysis of the genes mapped to SCZ and lipids implicated enrichment in “prion disease pathway” (Tables S94-97). Prion diseases are rare, neurodegenerative diseases caused by an infectious agent known as a prion. Pathway analysis for SCZ and TG also indicated “EPONFKB pathway” (Table S94), which is a pathway involving erythropoietin mediated neuroprotection through NF-kB. Pathway analysis for SCZ and HDL revealed “mRNA processing” (Table S95). This process involves the conversion of precursor messenger RNA into mature messenger RNA (mRNA). Pathway analysis for SCZ and WHR indicated “mRNA Processing” (Table S98).

Some of the GWAS summary statistics (for SCZ, lipids and smoking) include imputed data from the Haplotype Reference Consortium (HRC). FUMA does not include HRC-imputed data in the reference dataset. Thus, we checked if there were any SNPs identified with cond/conjFDR that were lost in the FUMA analyses and found that no SNPs were missing functional annotation.

Genomic SEM results

We used Genomic SEM to further analyze the relationship between SCZ and CVD. In particular, we investigated the effect of BMI and smoking initiation on the genetic relationship between SCZ and CAD. Based on the results above (r_g) (Table S101), BMI and smoking initiation are positively correlated with CAD, while SCZ demonstrates a negative genetic correlation with BMI and a positive genetic correlation with smoking initiation. Thus, we controlled for the effect of these CVD risk factors with both positive (smoking) and negative (BMI) associations with SCZ on the relationship between SCZ and CAD. We hypothesized that controlling for these CVD risk factors would modify the association between SCZ and CAD. In addition, we performed similar analyses with other CVD risk factors (one representative of each CVD risk category (e.g. blood pressure, lipids and T2D) that shared variants with mixed effect directions in SCZ and CAD (Table S101).

Multiple regression

Figure S17 shows the path diagrams for each of the models examined: The correlations between SCZ and CAD (Figure S17a-b) controlling for BMI were non-significant, similar to the bivariate correlations calculated with LDSR (Table S101). In particular, the path from SCZ to CAD (r_g=0.016, p=0.531, Figure S17a) and vice versa (r_g=0.017, p=0.531, Figure S17b) were close to zero. The correlations between SCZ and CAD (Figure S17c-d)

controlling for smoking initiation were also weak, yet significant (SCZ to CAD r_g=-0.055, p=0.041; CAD to SCZ r_g=-0.056, p=0.042). This indicates that controlling for the effect of smoking initiation and BMI have little to no impact on the relationship between SCZ and CAD. The analyses controlling for other CVD risk factors (lipid, SBP, T2D) did not significantly change the genetic correlation between SCZ and CAD (r_g<-0.018), p>0.528) (Figure S17c-i). The residual variance for CAD and SCZ was high (>0.83, number to the right in Figure S17), indicating that most of the variance in CAD and SCZ is not explained by the models.

Mediation

Figure S18 shows the mediation models for SCZ and CAD with CVD risk factors as mediators. The correlations produced from the total effect path c (Figure S18) are the same as the bivariate correlations produced through LDSR (Table S101). The indirect effect is the product of paths a and b (Figure S18). This results in a minor indirect effect of BMI (r_g=-0.035, p=5.016x10^{-9}) (Figure S18a) and smoking initiation (r_g=0.036, p=7.639x10^{-8} (Figure S18b). BMI as a mediator lowers the correlation between SCZ and CAD, while smoking as a mediator increases the correlation, although both correlations are still close to zero. The analyses with other CVD risk factors except for SBP (r_g=0.034, p=0.5694) also suggest a statistically significant, yet minor, indirect effect (HDL r_g=0.0167, p=4.439x10^{-3}; T2D r_g=0.0255, p=2.887x10^{-3}) on the relationship between SCZ and CAD (Figure S18c-e).

Multiple regression vs mediation analyses

The direct effect of SCZ on CAD and vice versa controlling for CVD risk factors (except for smoking initiation) was statistically insignificant in multiple regression analyses (Figure S17), while the indirect effects in the mediation models were statistically significant except for SBP

(Figure S18). This difference is probably due to the fact that tests for indirect effects have more power than the tests for direct effects (47).

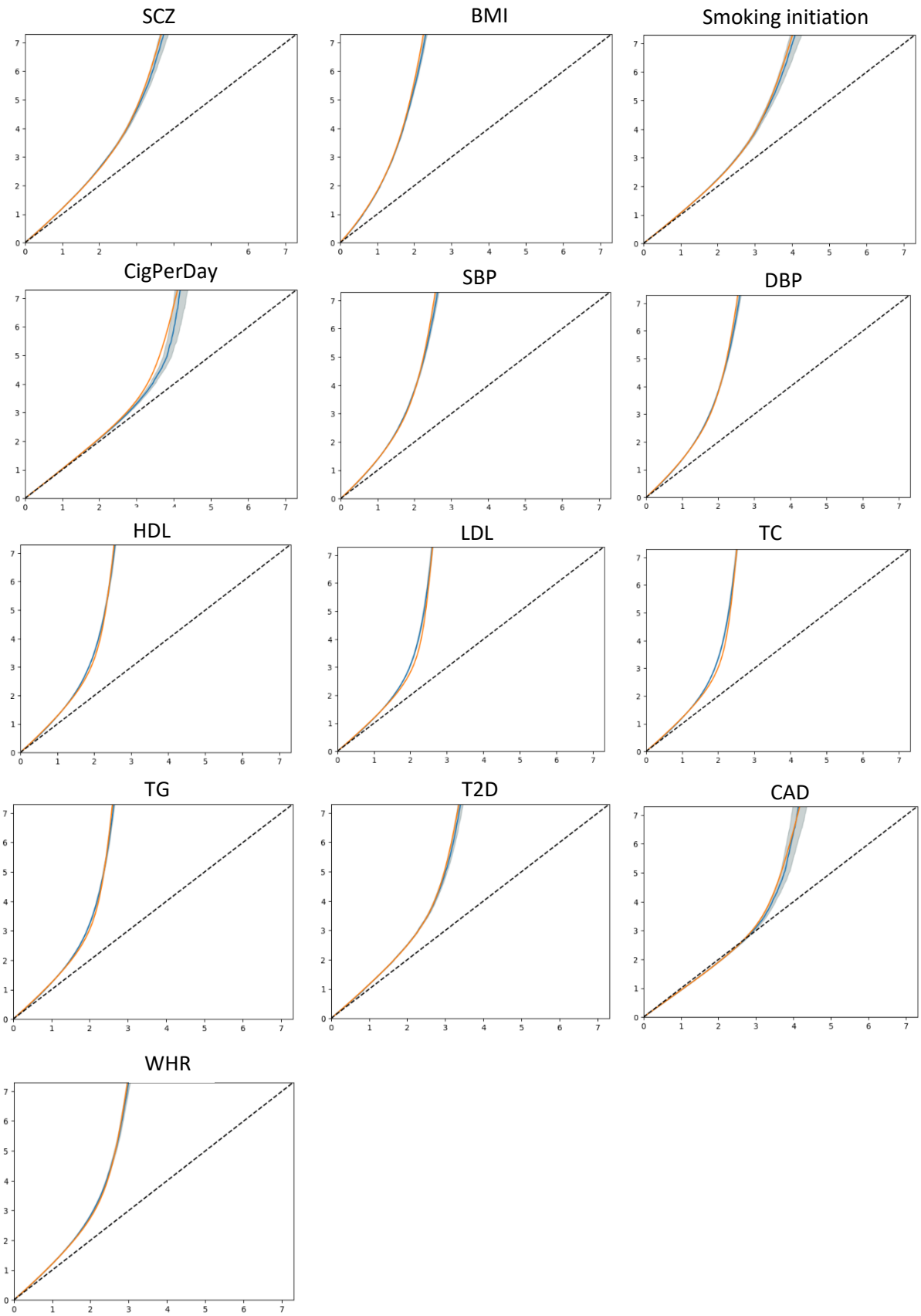


Figure S1. Univariate Q-Q plots for distribution of expected p-values under a null model (no SNPs associated with the phenotype) (x axis) versus observed p-values (y axis). Univariate Q-Q plots demonstrate that MiXeR-based predictions provide accurate estimates of the data Q-Q plots. Blue lines indicate p-values of SNPs observed in GWAS summary statistics with grey shading indicating 95% confidence interval. Orange lines indicate model predictions. The dashed line is the expected Q-Q plot under null (no SNPs associated with the phenotype). The vertical axes are limited to the genome-wide significance threshold of $p<5\times10^{-8}$, to highlight behavior of polygenic component. Points on the Q-Q plot are weighted according to LD structure, using n=64 iterations of random pruning at LD threshold r2=0.1.

A) SCZ

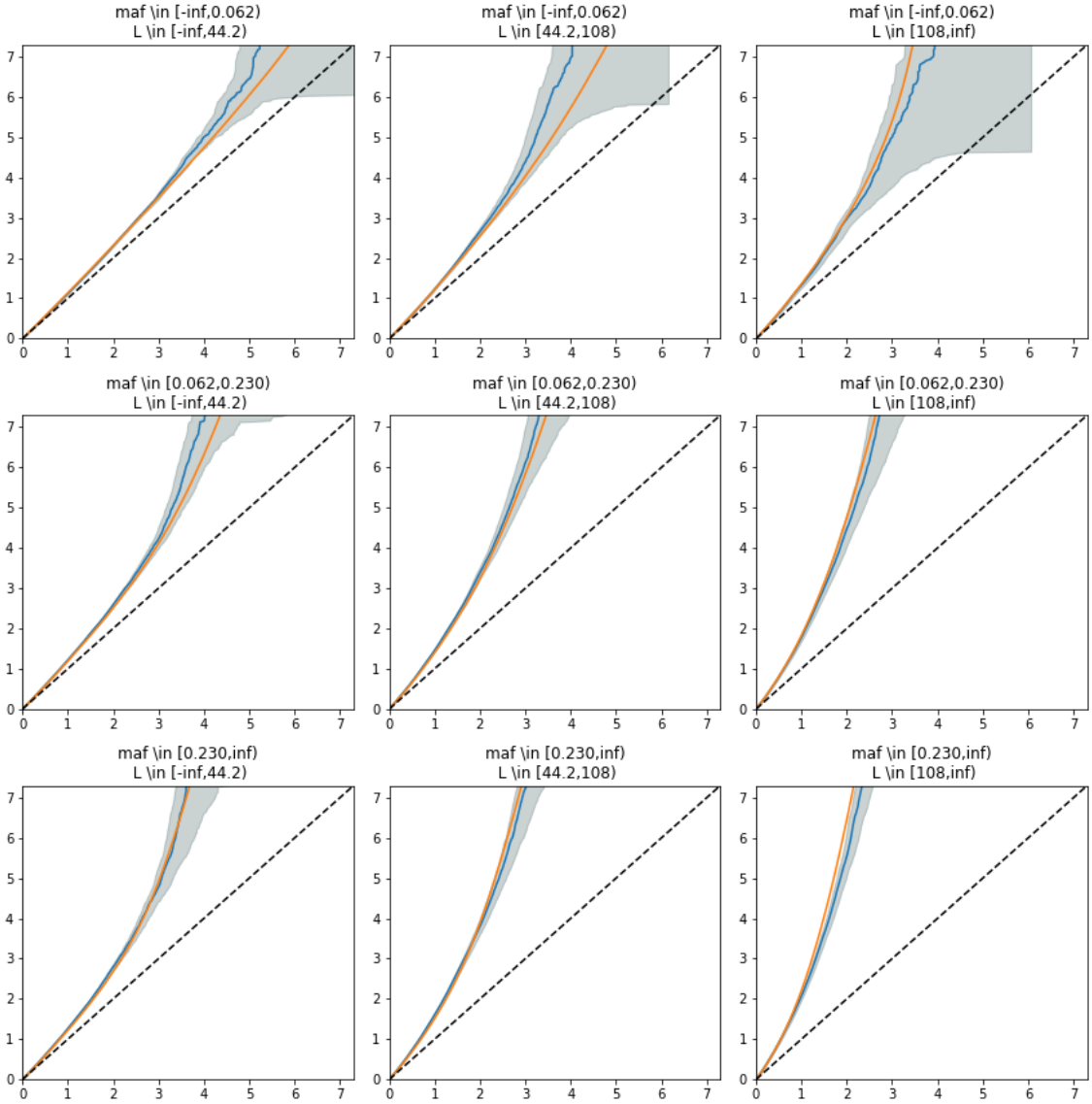


Figure S2. A) Univariate Q-Q plots for subsets of SNPs, showing observed schizophrenia (SCZ) GWAS p-values (in blue) and model prediction (in orange). Grey shading indicating 95% confidence interval. All SNPs were partitioned into 9 groups according to minor allele frequency (MAF) and total linkage disequilibrium (LD) score. Observed Q-Q plots show a stronger GWAS signal for SNPs with higher MAF and larger total LD score.

B) BMI

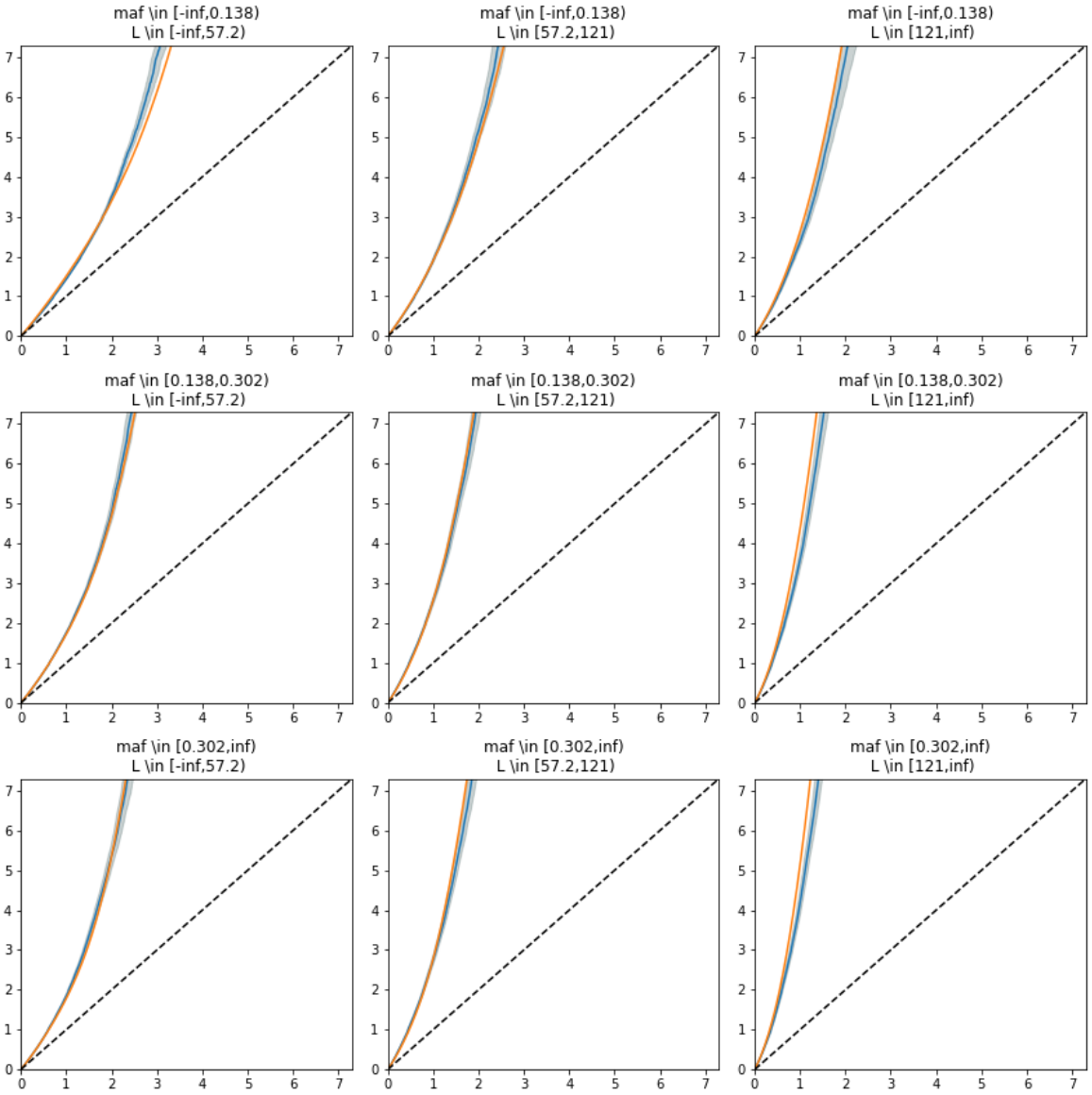


Figure S2. B) Univariate Q-Q plots for subsets of SNPs, showing observed body mass index (BMI) GWAS p-values (in blue) and model prediction (in orange). Grey shading indicating 95% confidence interval. All SNPs were partitioned into 9 groups according to minor allele frequency (MAF) and total linkage disequilibrium (LD) score. Observed Q-Q plots show a stronger GWAS signal for SNPs with higher MAF and larger total LD score.

C) Smoking initiation

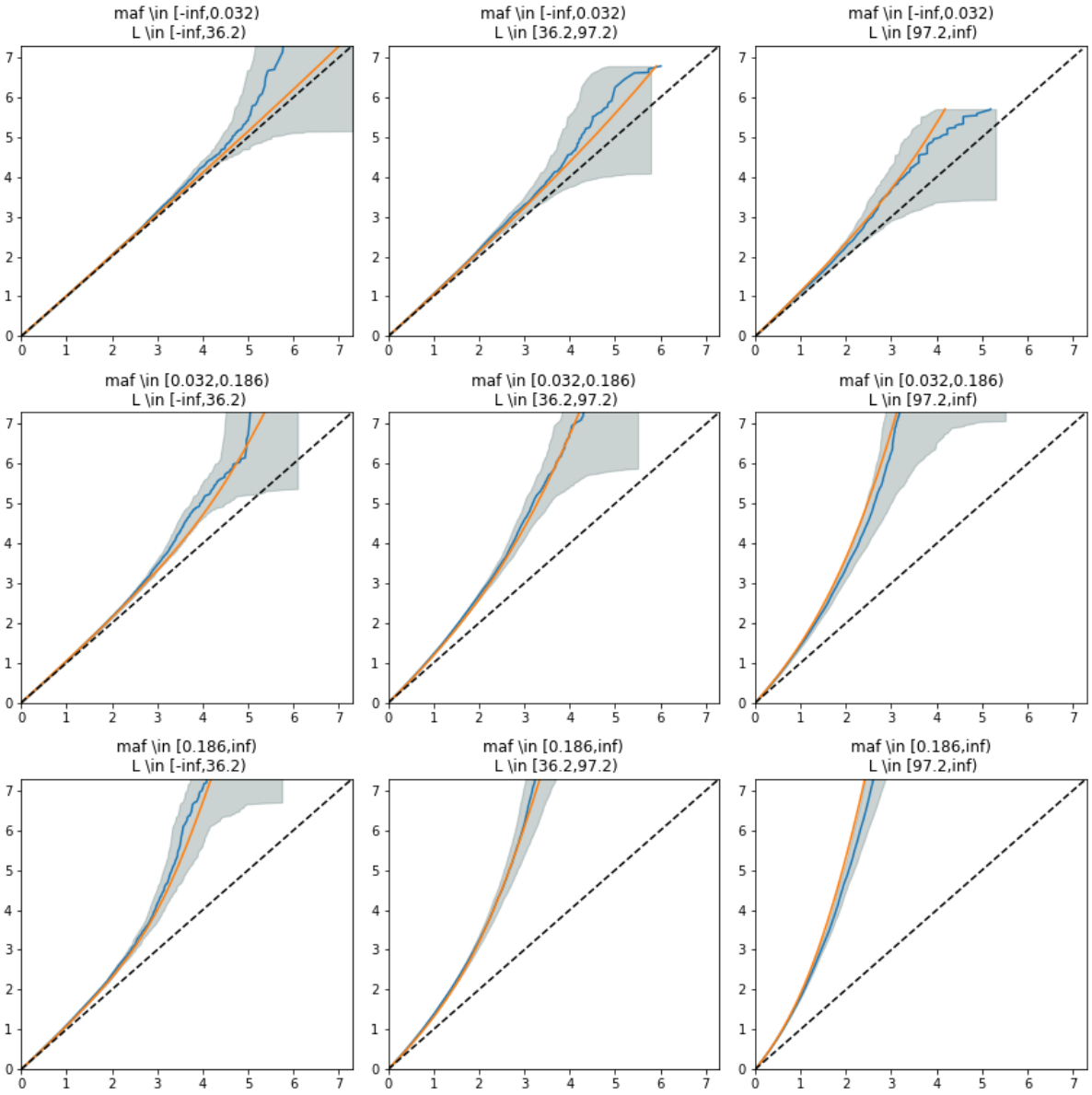


Figure S2. C) Univariate Q-Q plots for subsets of SNPs, showing observed smoking initiation GWAS p-values (in blue) and model prediction (in orange). Grey shading indicating 95% confidence interval. All SNPs were partitioned into 9 groups according to minor allele frequency (MAF) and total linkage disequilibrium (LD) score. Observed Q-Q plots show a stronger GWAS signal for SNPs with higher MAF and larger total LD score.

D) CigPerDay

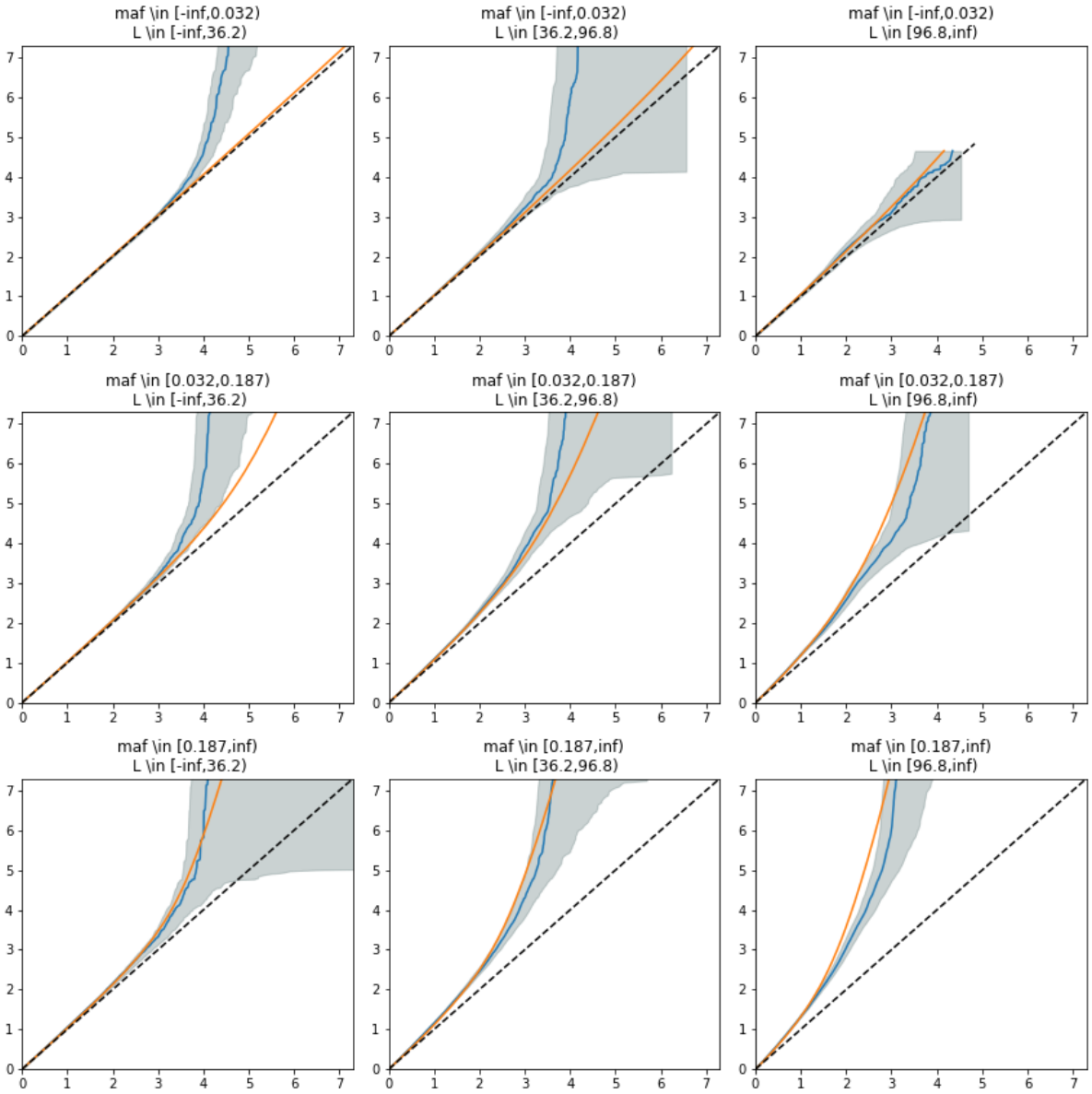


Figure S2. D) Univariate Q-Q plots for subsets of SNPs, showing observed cigarettes per day (CigPerDay) GWAS p-values (in blue) and model prediction (in orange). Grey shading indicating 95% confidence interval. All SNPs were partitioned into 9 groups according to minor allele frequency (MAF) and total linkage disequilibrium (LD) score. Observed Q-Q plots show a stronger GWAS signal for SNPs with higher MAF and larger total LD score.

E) SBP

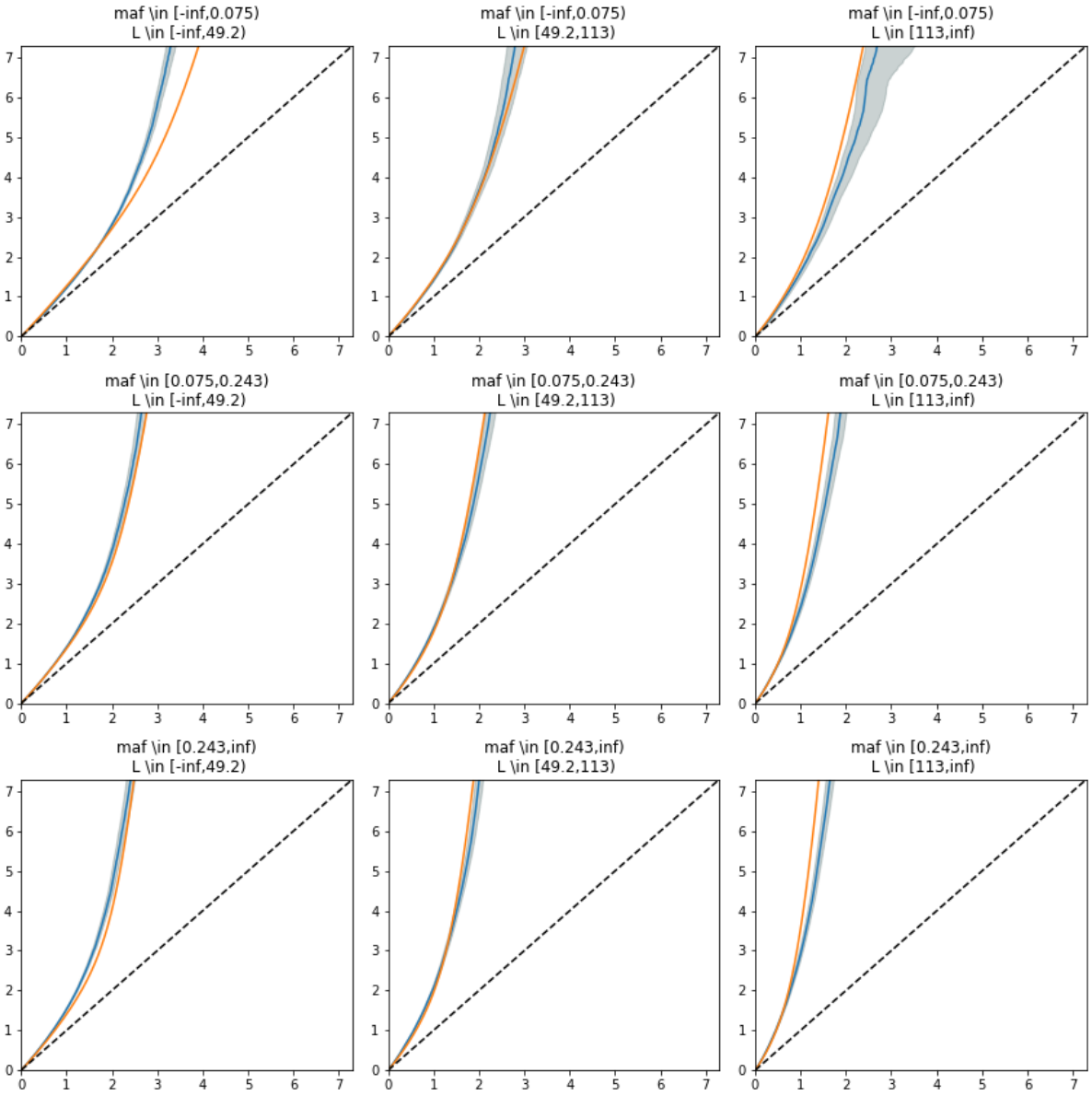


Figure S2. E) Univariate Q-Q plots for subsets of SNPs, showing observed systolic blood pressure (SBP) GWAS p-values (in blue) and model prediction (in orange). Grey shading indicating 95% confidence interval. All SNPs were partitioned into 9 groups according to minor allele frequency (MAF) and total linkage disequilibrium (LD) score. Observed Q-Q plots show a stronger GWAS signal for SNPs with higher MAF and larger total LD score.

F) DBP

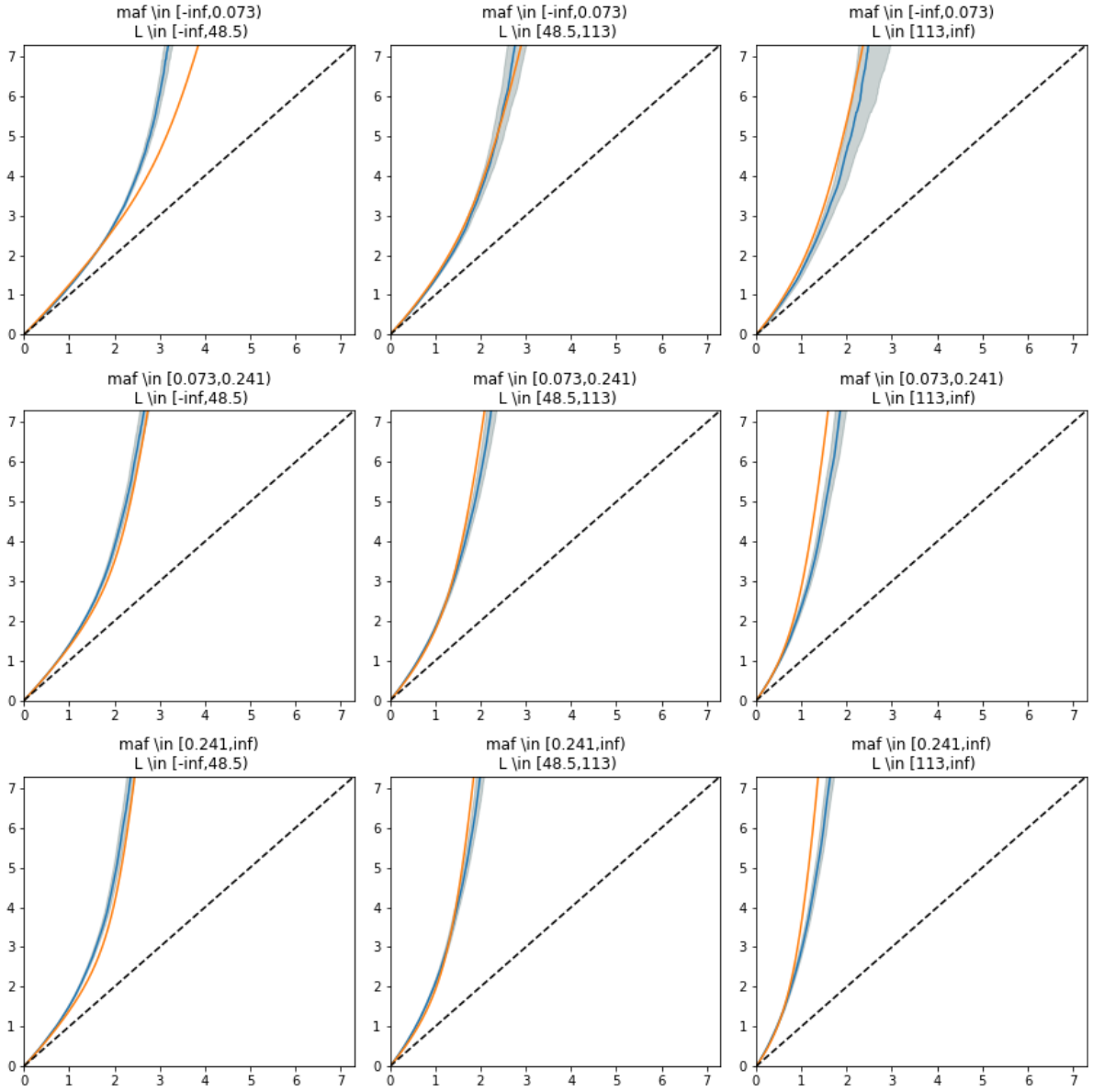


Figure S2. F) Univariate Q-Q plots for subsets of SNPs, showing observed diastolic blood pressure (DBP) GWAS p-values (in blue) and model prediction (in orange). Grey shading indicating 95% confidence interval. All SNPs were partitioned into 9 groups according to minor allele frequency (MAF) and total linkage disequilibrium (LD) score. Observed Q-Q plots show a stronger GWAS signal for SNPs with higher MAF and larger total LD score.

G) HDL

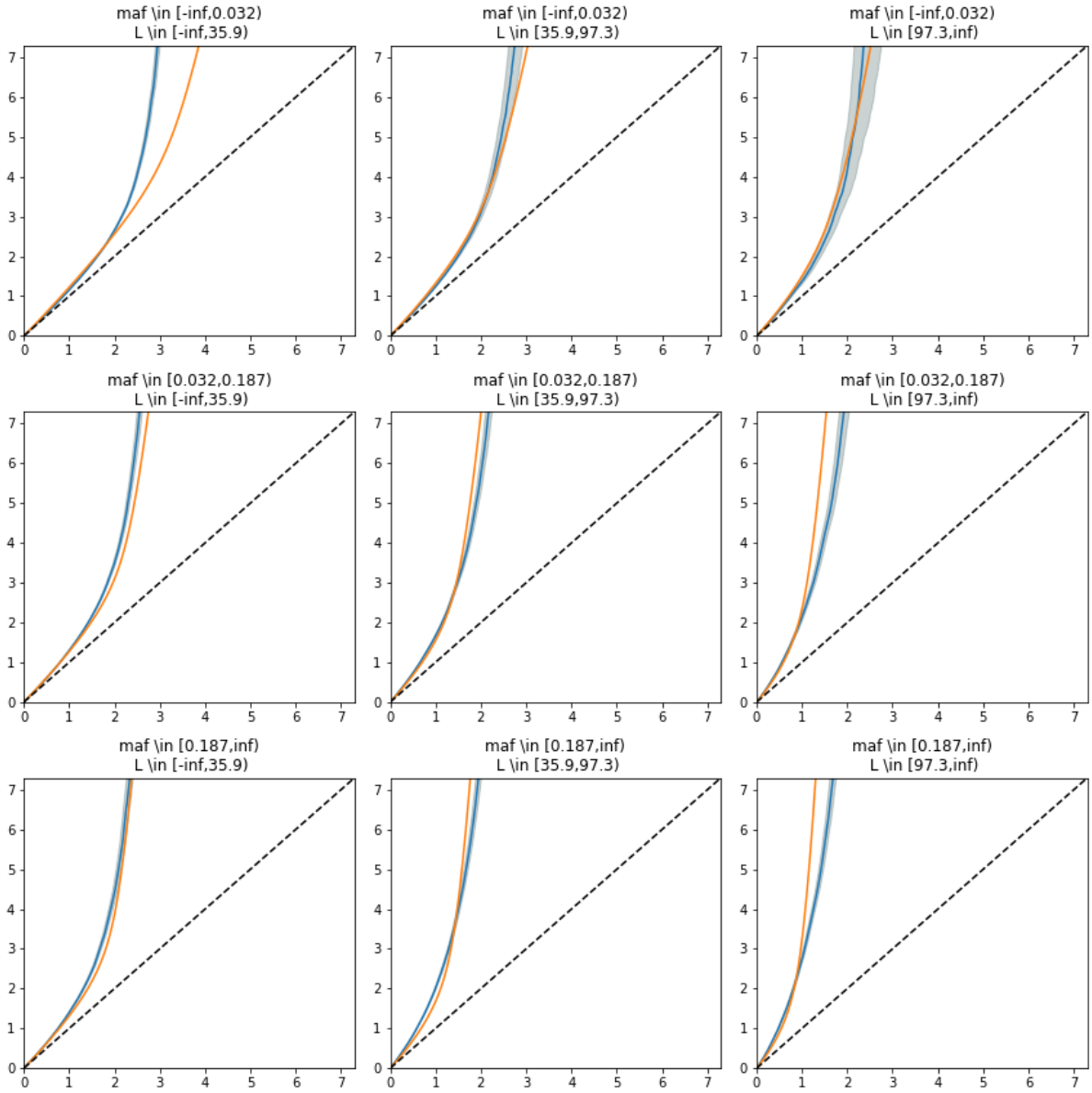


Figure S2. G) Univariate Q-Q plots for subsets of SNPs, showing observed high-density lipoprotein cholesterol (HDL) GWAS p-values (in blue) and model prediction (in orange). Grey shading indicating 95% confidence interval. All SNPs were partitioned into 9 groups according to minor allele frequency (MAF) and total linkage disequilibrium (LD) score. Observed Q-Q plots show a stronger GWAS signal for SNPs with higher MAF and larger total LD score.

H) LDL

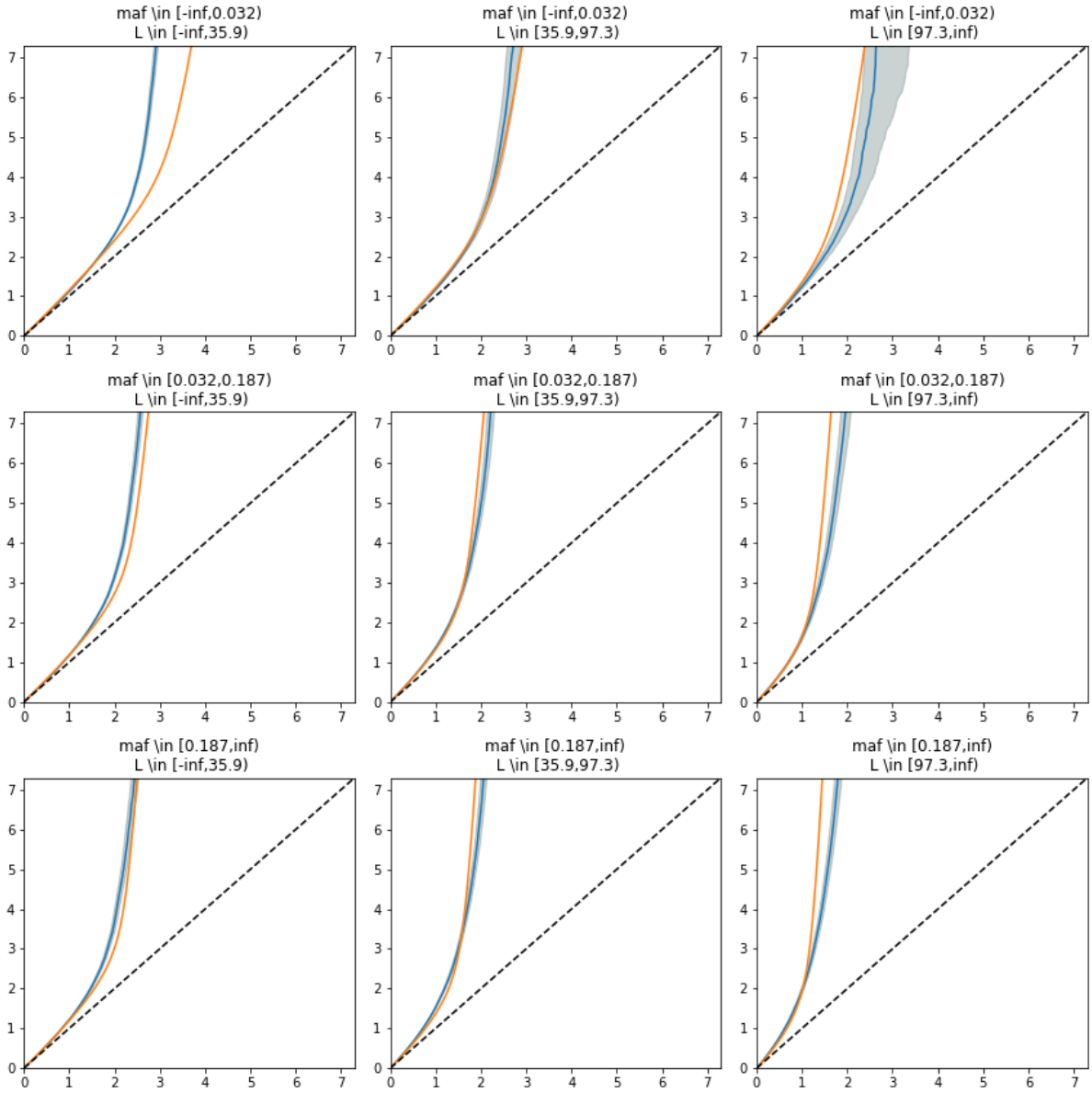


Figure S2. H) Univariate Q-Q plots for subsets of SNPs, showing observed low-density lipoprotein cholesterol (LDL) GWAS p-values (in blue) and model prediction (in orange). Grey shading indicating 95% confidence interval. All SNPs were partitioned into 9 groups according to minor allele frequency (MAF) and total linkage disequilibrium (LD) score. Observed Q-Q plots show a stronger GWAS signal for SNPs with higher MAF and larger total LD score.

I) TC

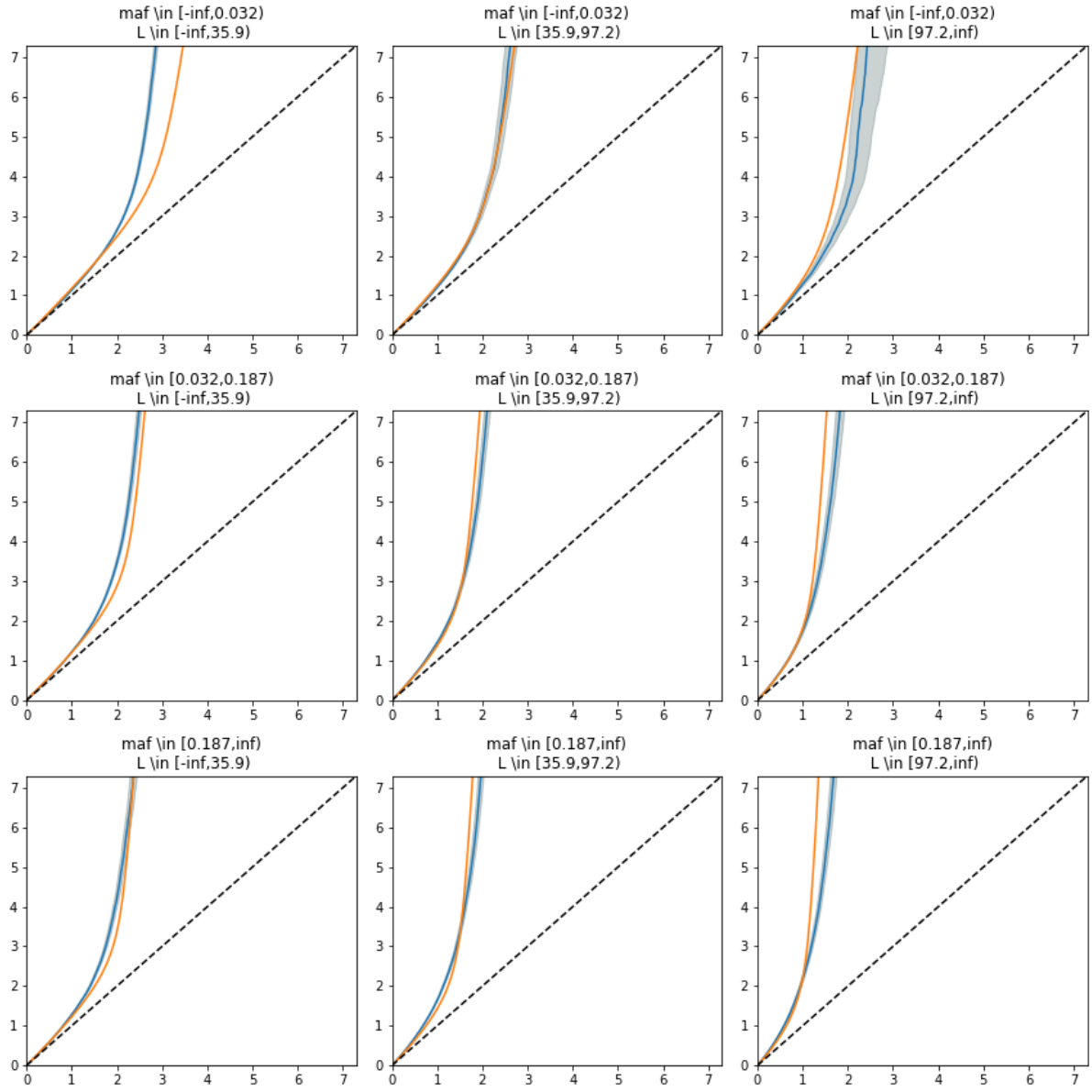


Figure S2. I) Univariate Q-Q plots for subsets of SNPs, showing observed total cholesterol (TC) GWAS p-values (in blue) and model prediction (in orange). Grey shading indicating 95% confidence interval. All SNPs were partitioned into 9 groups according to minor allele frequency (MAF) and total linkage disequilibrium (LD) score. Observed Q-Q plots show a stronger GWAS signal for SNPs with higher MAF and larger total LD score.

J) TG

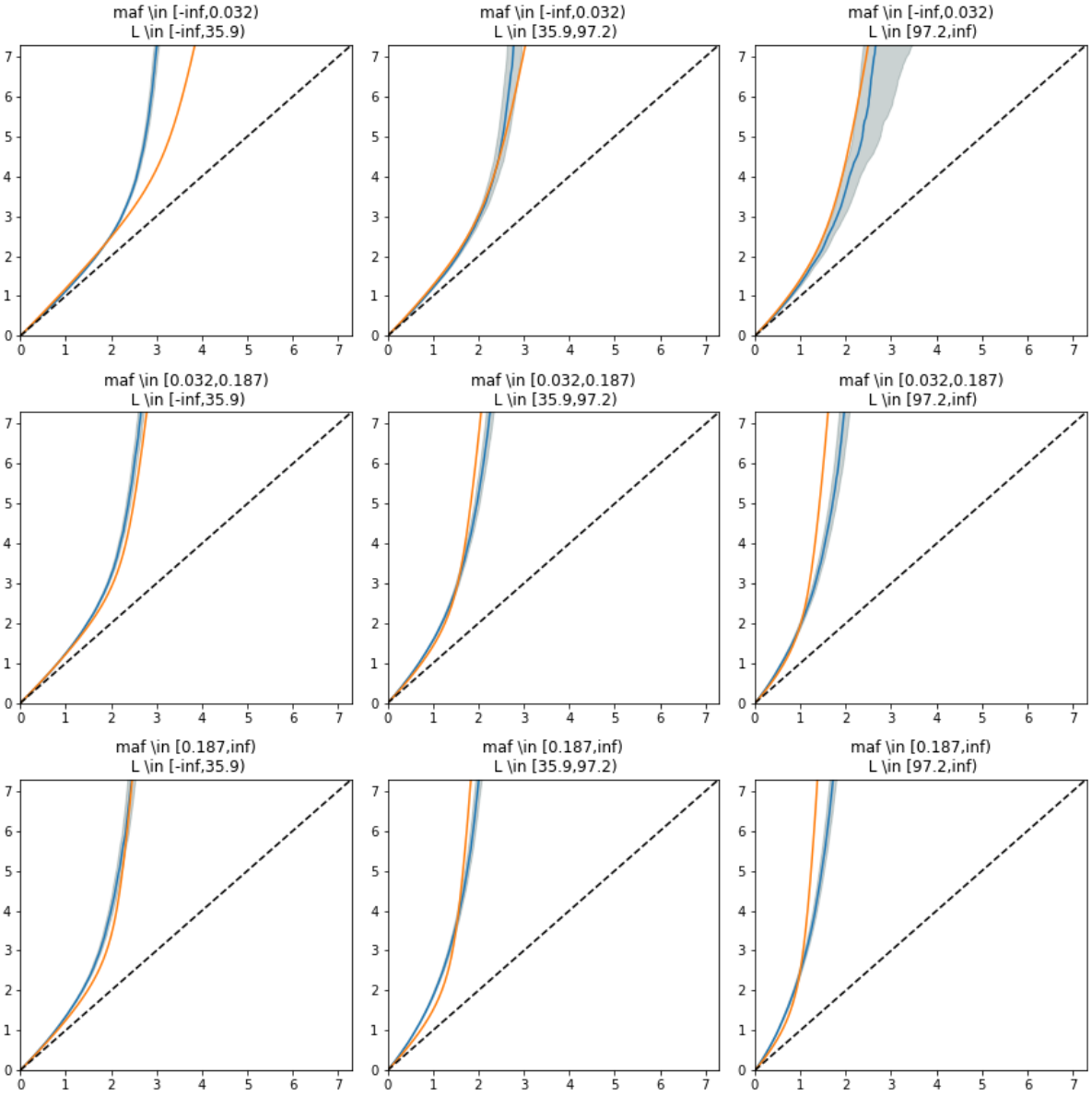


Figure S2. J) Univariate Q-Q plots for subsets of SNPs, showing observed triglycerides (TG) GWAS p-values (in blue) and model prediction (in orange). Grey shading indicating 95% confidence interval. All SNPs were partitioned into 9 groups according to minor allele frequency (MAF) and total linkage disequilibrium (LD) score. Observed Q-Q plots show a stronger GWAS signal for SNPs with higher MAF and larger total LD score.

K) T2D

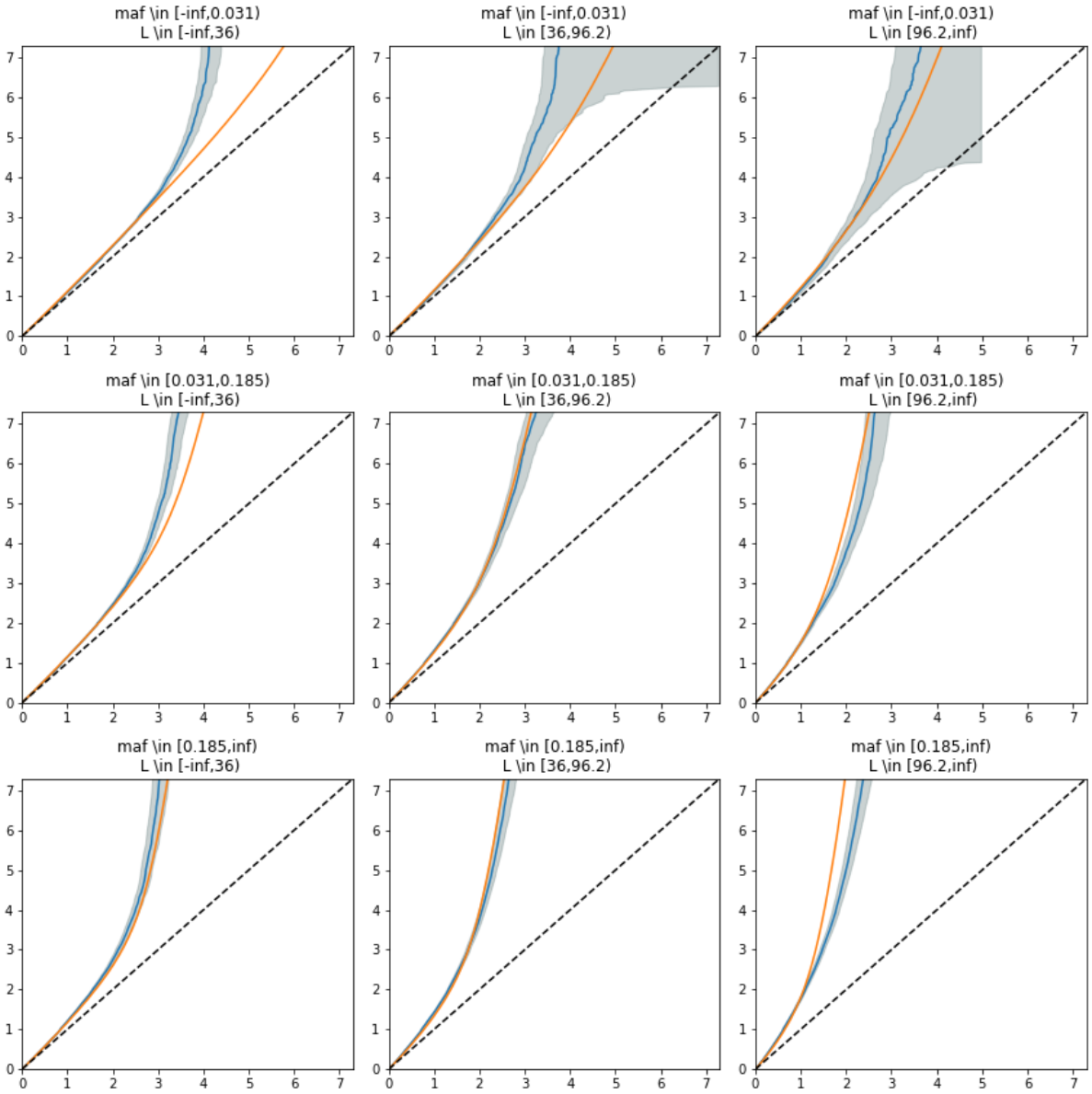


Figure S2. K) Univariate Q-Q plots for subsets of SNPs, showing observed type 2 diabetes (T2D) GWAS p-values (in blue) and model prediction (in orange). Grey shading indicating 95% confidence interval. All SNPs were partitioned into 9 groups according to minor allele frequency (MAF) and total linkage disequilibrium (LD) score. Observed Q-Q plots show a stronger GWAS signal for SNPs with higher MAF and larger total LD score.

L) CAD

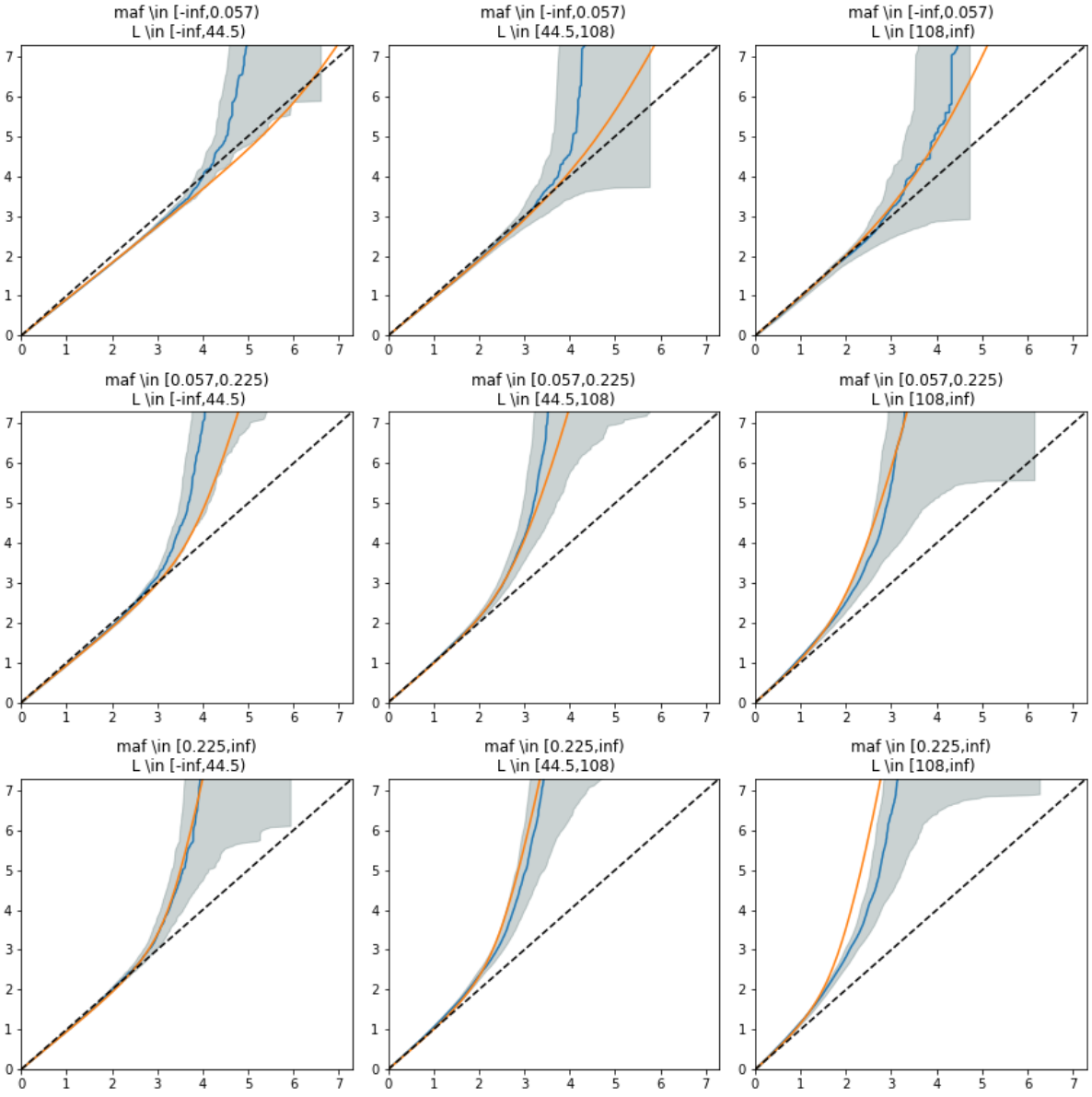


Figure S2. L) Univariate Q-Q plots for subsets of SNPs, showing observed coronary artery disease (CAD) GWAS p-values (in blue) and model prediction (in orange). Grey shading indicating 95% confidence interval. All SNPs were partitioned into 9 groups according to minor allele frequency (MAF) and total linkage disequilibrium (LD) score. Observed Q-Q plots show a stronger GWAS signal for SNPs with higher MAF and larger total LD score.

M) WHR

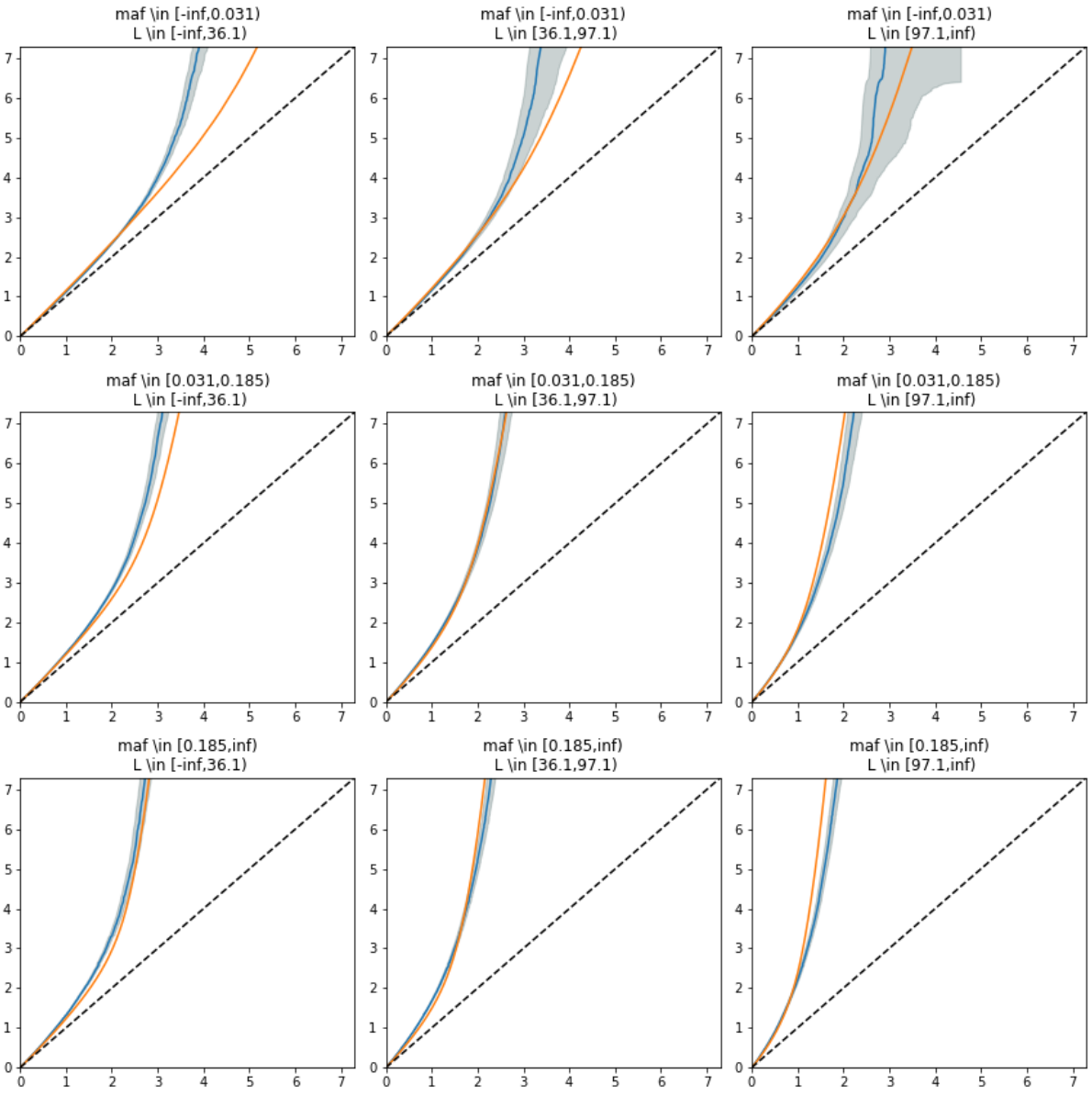


Figure S2. M) Univariate Q-Q plots for subsets of SNPs, showing observed waist-to-hip ratio (WHR) GWAS p-values (in blue) and model prediction (in orange). Grey shading indicating 95% confidence interval. All SNPs were partitioned into 9 groups according to minor allele frequency (MAF) and total linkage disequilibrium (LD) score. Observed Q-Q plots show a stronger GWAS signal for SNPs with higher MAF and larger total LD score.

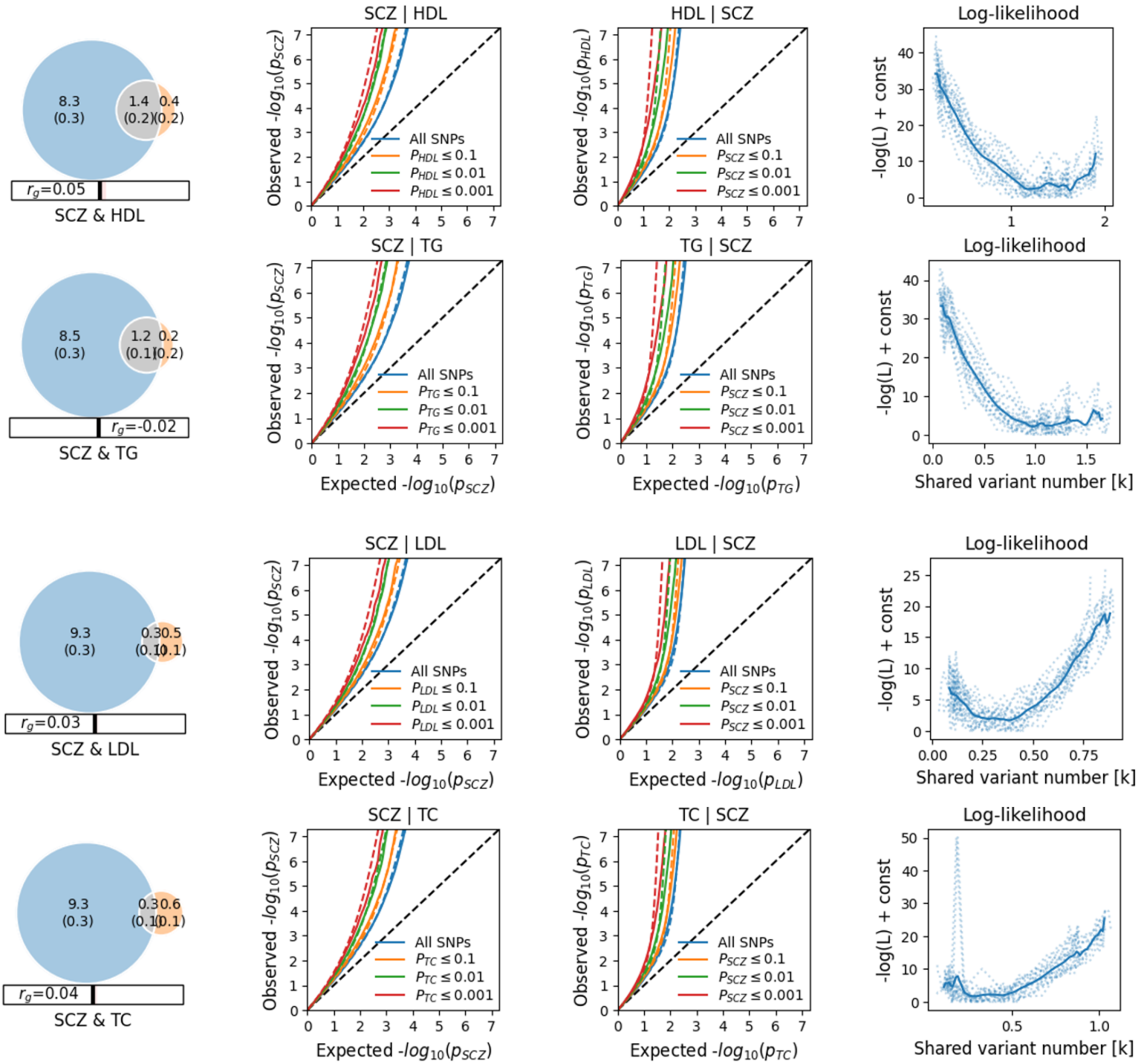


Figure S3. Venn Diagrams, conditional Q-Q plots, and negative log-likelihood plot, respectively. Venn diagrams of shared and unique trait-influencing variants, showing polygenic overlap (gray) between schizophrenia (SCZ) (blue) and lipids (HDL, TG, LDL, and TC) (orange). The numbers in the Venn diagram indicate the estimated quantity of trait-influencing variants (in thousands), explaining 90% of SNP heritability in each phenotype, followed by standard error. Conditional Q–Q plots of observed versus expected $-\log_{10}$ p-values in the primary trait as a function of significance of association with a secondary trait at the level of $p < 0.1$, $p < 0.01$, $p < 0.001$. Blue line indicates all SNPs. Dotted lines in blue, orange, green, and red indicate model predictions for each stratum. Black dotted line is the expected Q–Q plot under null hypothesis. Negative log-likelihood plot: minus log-likelihood calculated for the bivariate model as a function of π parameter. The remaining parameters of the model were constrained to their fitted values. Figure generated from MiXeR.

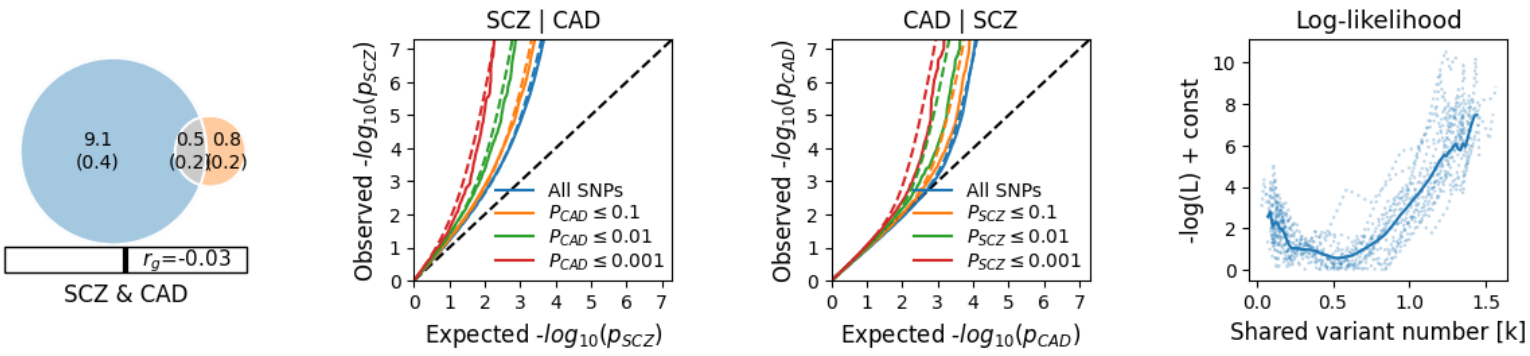


Figure S4. Venn Diagrams, conditional Q-Q plots, and negative log-likelihood plot, respectively. Venn diagrams of shared and unique trait-influencing variants, showing polygenic overlap (gray) between schizophrenia (SCZ) (blue) and coronary artery disease (CAD) (orange). The numbers in the Venn diagram indicate the estimated quantity of trait-influencing variants (in thousands), followed by standard error. Appearance of the Q-Q plot and negative log-likelihood plot are described below Figure S3. Figure generated from MiXeR.

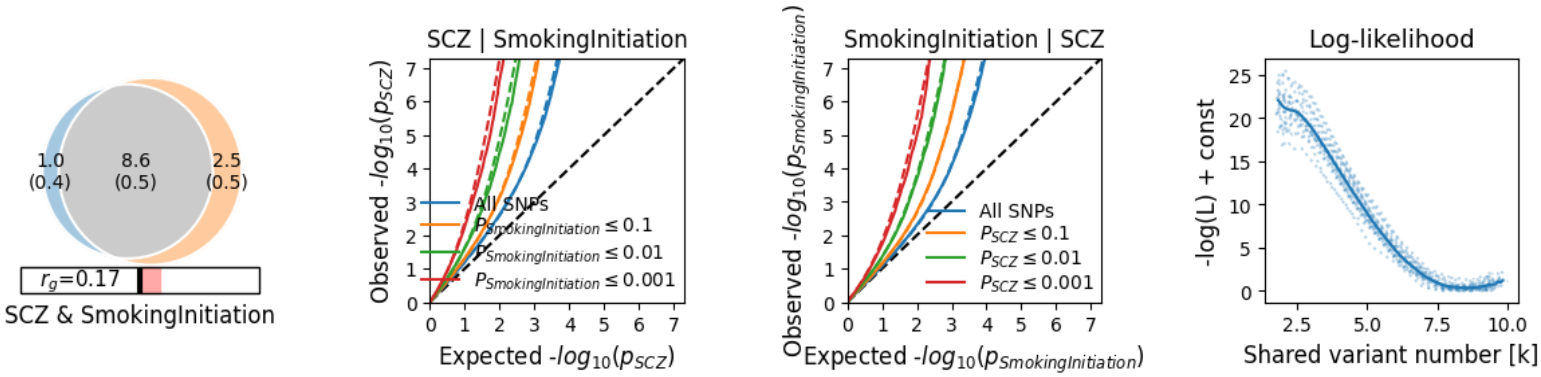


Figure S5. Venn Diagrams, conditional Q-Q plots, and negative log-likelihood plot, respectively. *Venn diagrams* of shared and unique trait-influencing variants, showing polygenic overlap (gray) between schizophrenia (SCZ) (blue) and smoking initiation (orange). The numbers in the Venn diagram indicate the estimated quantity of trait-influencing variants (in thousands), explaining 90% of SNP heritability in each phenotype, followed by standard error. Appearance of the Q-Q plot and negative log-likelihood plot are described below Figure S3. Figure generated from MiXeR.

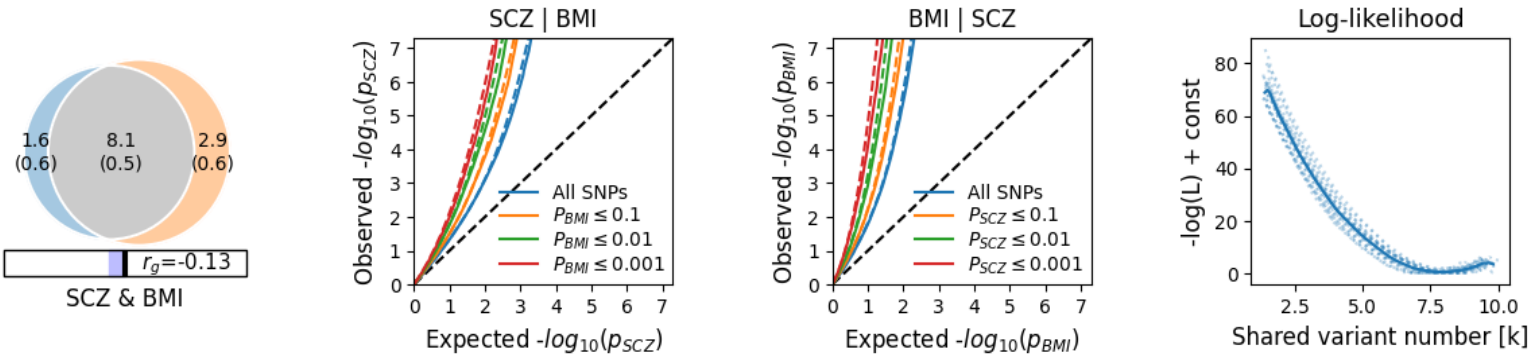


Figure S6. Venn Diagrams, conditional Q-Q plots, and negative log-likelihood plot, respectively. Venn diagrams of shared and unique trait-influencing variants, showing polygenic overlap (gray) between schizophrenia (SCZ) (blue) and body mass index (BMI) (orange). The numbers in the Venn diagram indicate the estimated quantity of trait-influencing variants (in thousands), explaining 90% of SNP heritability in each phenotype, followed by standard error. Appearance of the Q-Q plot and negative log-likelihood plot are described below Figure S3. Figure generated from MiXeR.

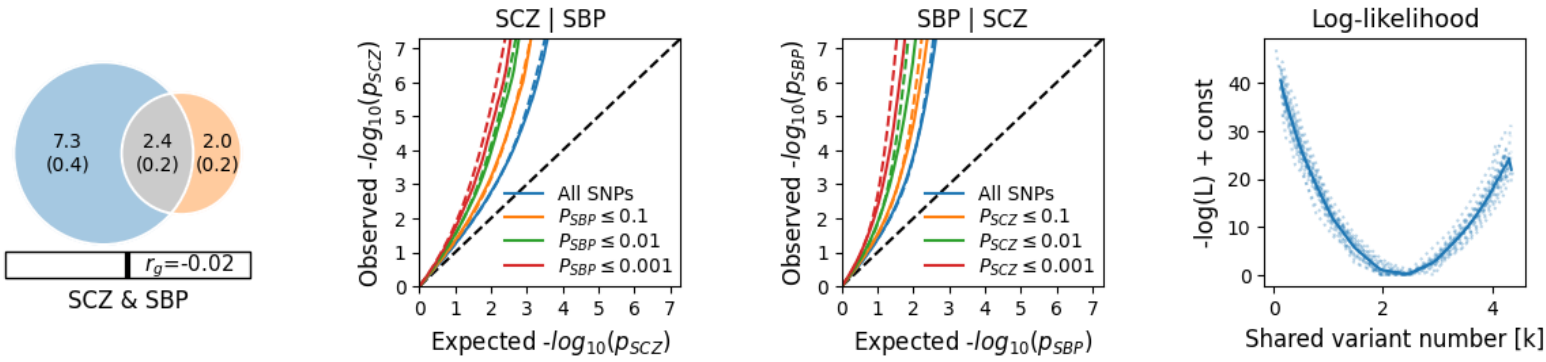


Figure S7. Venn Diagrams, conditional Q-Q plots, and negative log-likelihood plot, respectively. Venn diagrams of shared and unique trait-influencing variants, showing polygenic overlap (gray) between schizophrenia (SCZ) (blue) and systolic blood pressure (SBP) (orange). The numbers in the Venn diagram indicate the estimated quantity of trait-influencing variants (in thousands), followed by standard error. Appearance of the Q-Q plot and negative log-likelihood plot are described below Figure S3. Figure generated from MiXeR.

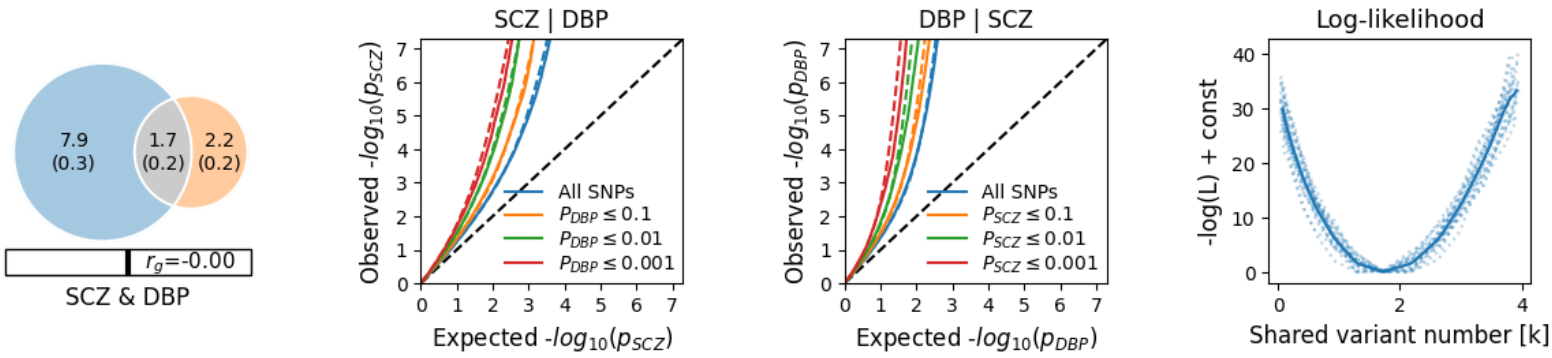


Figure S8. Venn Diagrams, conditional Q-Q plots, and negative log-likelihood plot, respectively. Venn diagrams of shared and unique trait-influencing variants, showing polygenic overlap (gray) between schizophrenia (SCZ) (blue) and diastolic blood pressure (DBP) (orange). The numbers in the Venn diagram indicate the estimated quantity of trait-influencing variants (in thousands), followed by standard error. Appearance of the Q-Q plot and negative log-likelihood plot are described below Figure S3. Figure generated from MiXeR.

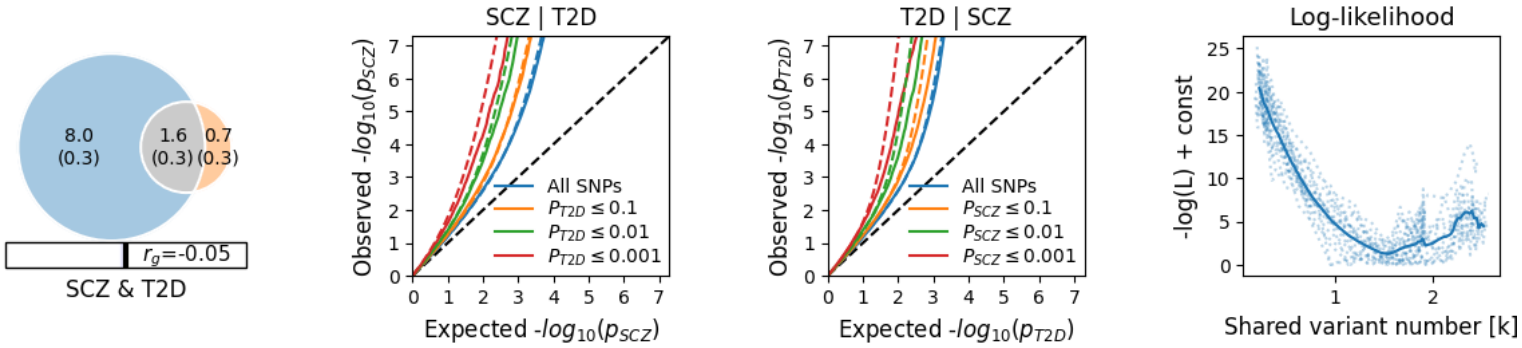


Figure S9. Venn Diagrams, conditional Q-Q plots, and negative log-likelihood plot, respectively. Venn diagrams of shared and unique trait-influencing variants, showing polygenic overlap (gray) between schizophrenia (SCZ) (blue) and type 2 diabetes (T2D) (orange). The numbers in the Venn diagram indicate the estimated quantity of trait-influencing variants (in thousands), followed by standard error. Appearance of the Q-Q plot and negative log-likelihood plot are described below Figure S3. Figure generated from MiXeR.

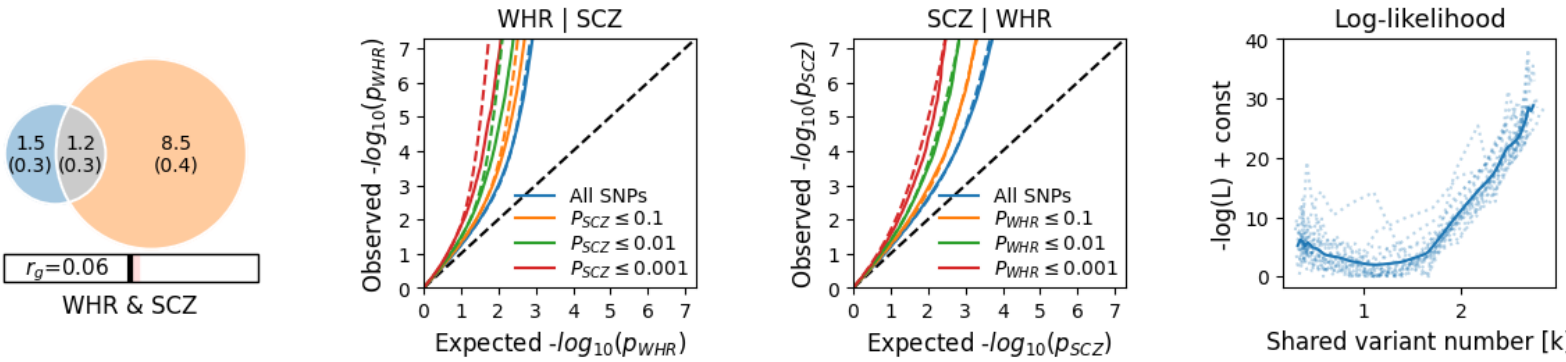


Figure S10. Venn Diagrams, conditional Q-Q plots, and negative log-likelihood plot, respectively. Venn diagrams of shared and unique trait-influencing variants, showing polygenic overlap (gray) between schizophrenia (SCZ) (blue) and waist-to-hip ratio (WHR) adjusted for BMI (orange). The numbers in the Venn diagram indicate the estimated quantity of trait-influencing variants (in thousands), followed by standard error. Appearance of the Q-Q plot and negative log-likelihood plot are described below Figure S3. Figure generated from MiXeR.

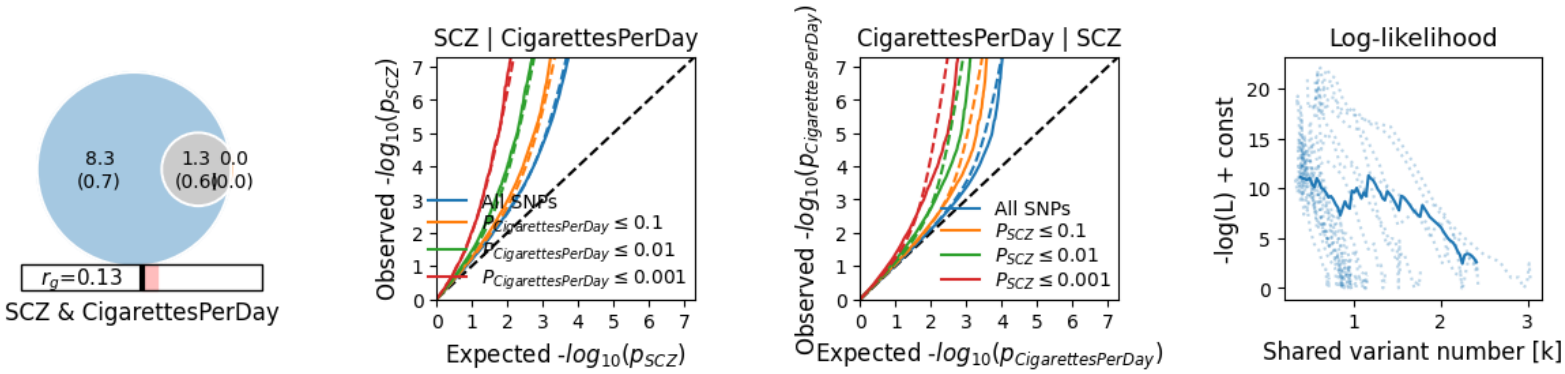
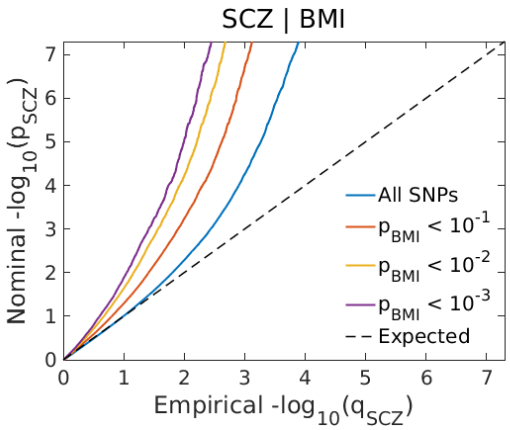
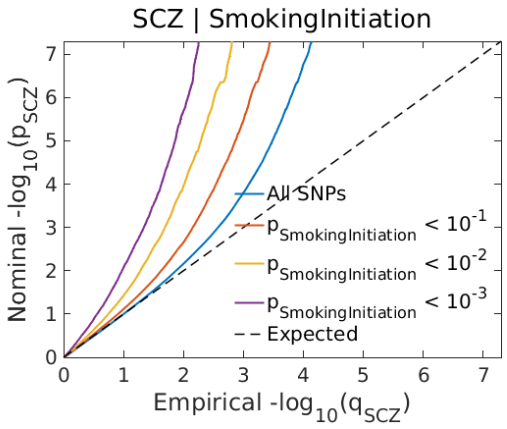
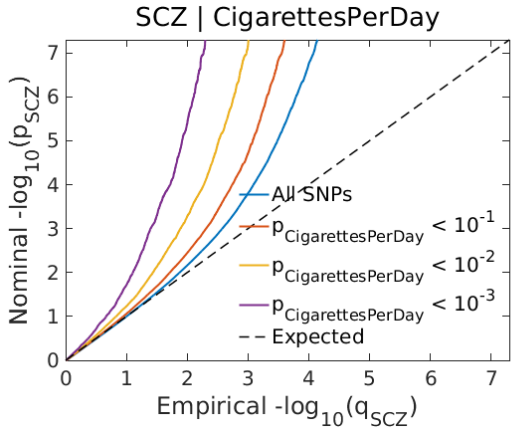
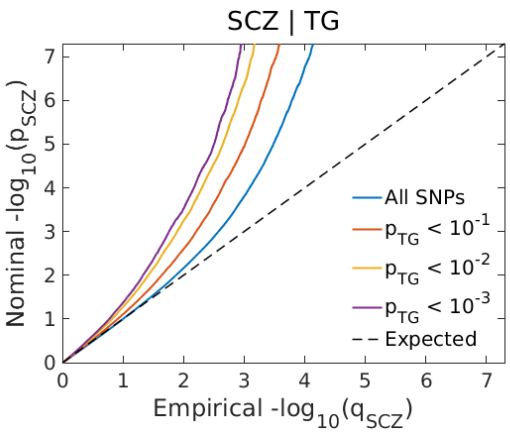


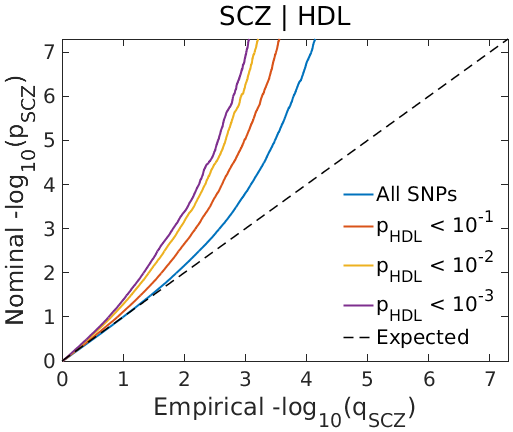
Figure S11. Venn Diagrams, conditional Q-Q plots, and negative log-likelihood plot, respectively. Venn diagrams of shared and unique trait-influencing variants, showing polygenic overlap (gray) between schizophrenia (SCZ) (blue) and Cigarettes per day (orange). The numbers in the Venn diagram indicate the estimated quantity of trait-influencing variants (in thousands), explaining 90% of SNP heritability in each phenotype, followed by standard error. Appearance of the Q-Q plot and negative log-likelihood plot are described below Figure S3. Figure generated from MiXeR.

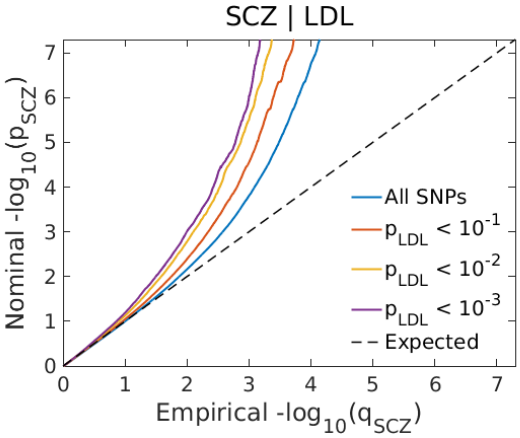

SCZ | BMI
Nominal -log10(pSCZ)
Empirical -log10(qSCZ)
All SNPs
pBMI < 10-1
pBMI < 10-2
pBMI < 10-3
Expected

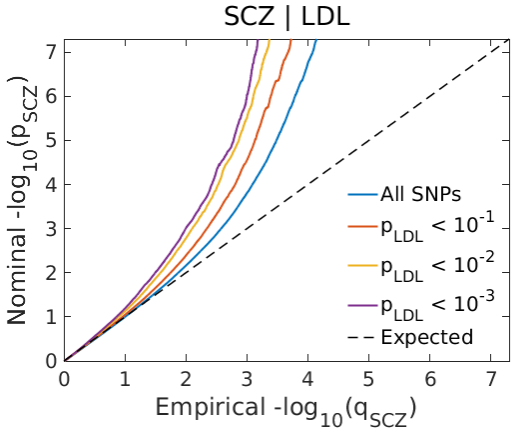

SCZ | SmokingInitiation
Nominal -log10(pSCZ)
Empirical -log10(qSCZ)
All SNPs
pSmokingInitiation < 10-1
pSmokingInitiation < 10-2
pSmokingInitiation < 10-3
Expected

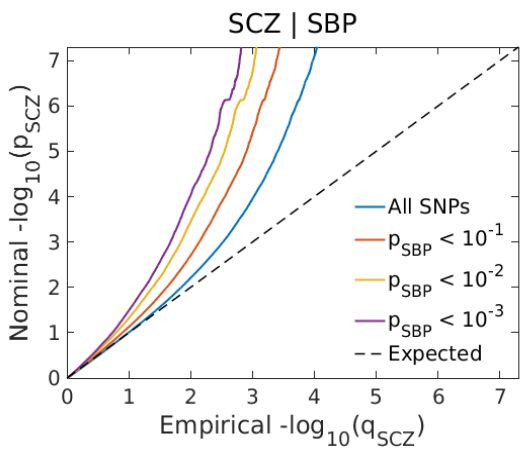

SCZ | CigarettesPerDay
Nominal -log10(pSCZ)
Empirical -log10(qSCZ)
All SNPs
pCigarettesPerDay < 10-1
pCigarettesPerDay < 10-2
pCigarettesPerDay < 10-3
Expected


SCZ | TG
Nominal -log10(pSCZ)
Empirical -log10(qSCZ)
All SNPs
pTG < 10-1
pTG < 10-2
pTG < 10-3
Expected


SCZ | HDL
Nominal -log10(pSCZ)
Empirical -log10(qSCZ)
All SNPs
pHDL < 10-1
pHDL < 10-2
pHDL < 10-3
Expected


SCZ | LDL
Nominal -log10(pSCZ)
Empirical -log10(qSCZ)
All SNPs
pLDL < 10-1
pLDL < 10-2
pLDL < 10-3
Expected


SCZ | LDL
Nominal -log10(pSCZ)
Empirical -log10(qSCZ)
All SNPs
pLDL < 10-1
pLDL < 10-2
pLDL < 10-3
Expected


SCZ | SBP
Nominal -log10(pSCZ)
Empirical -log10(qSCZ)
All SNPs
pSBP < 10-1
pSBP < 10-2
pSBP < 10-3
Expected

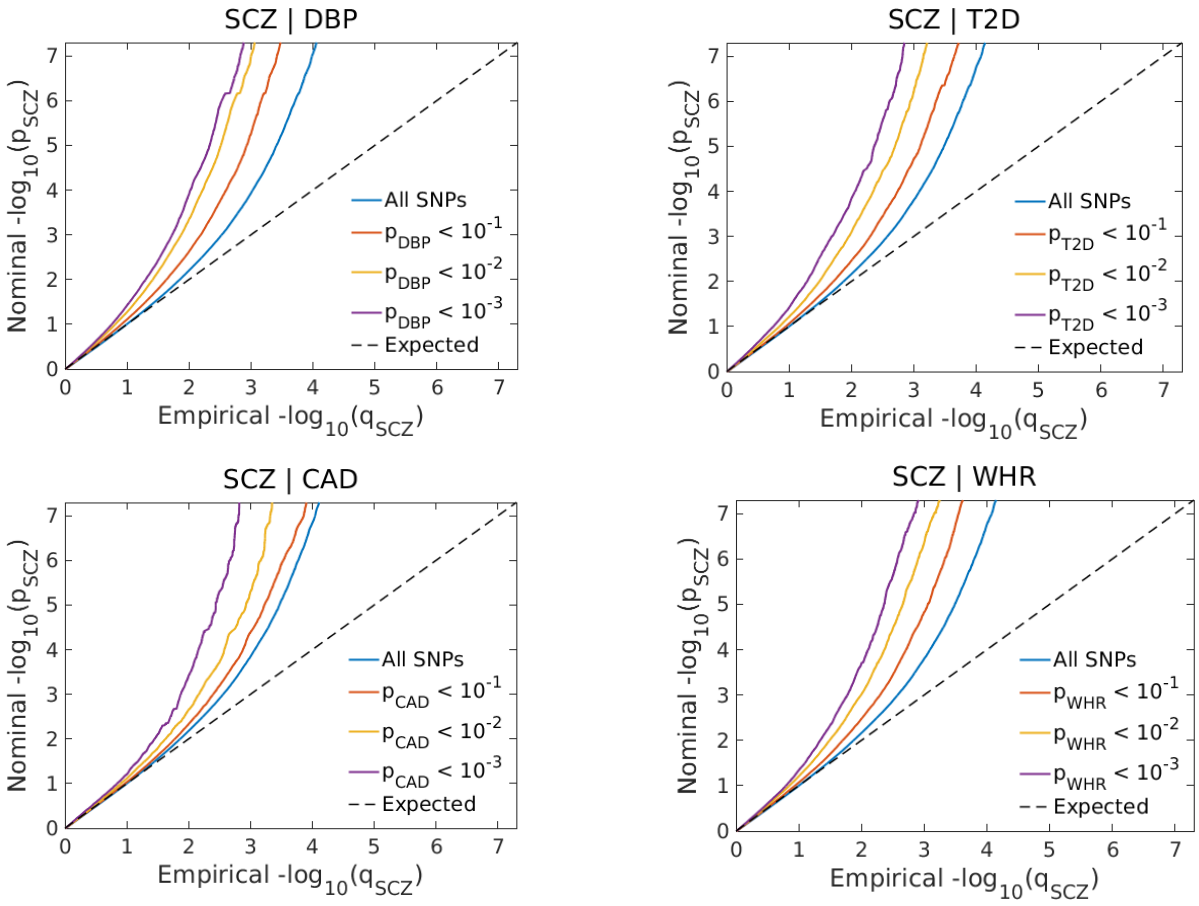
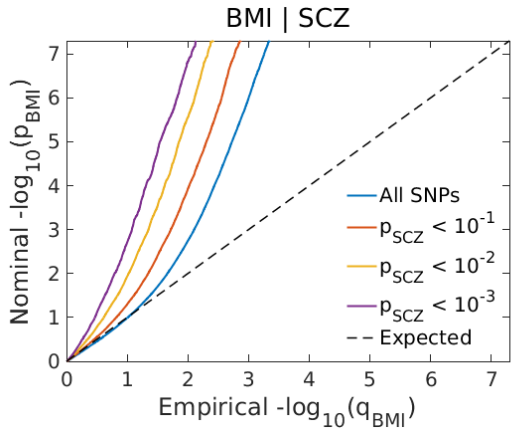
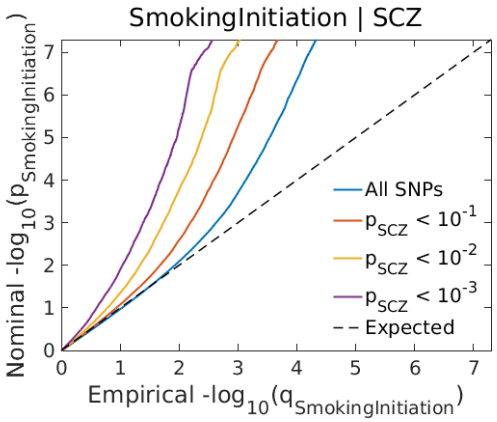
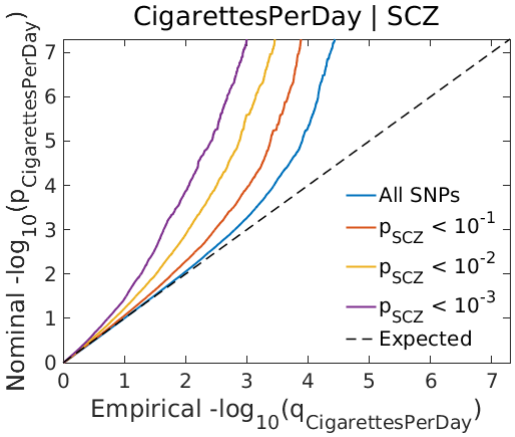
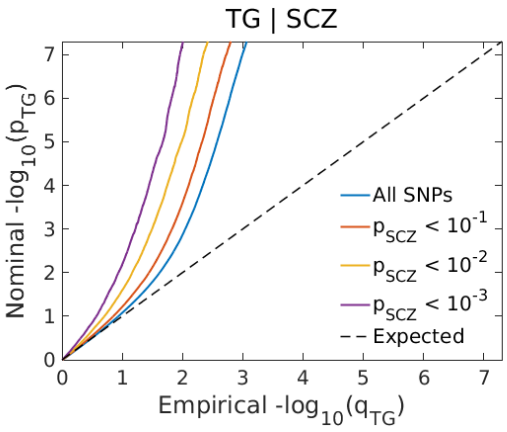


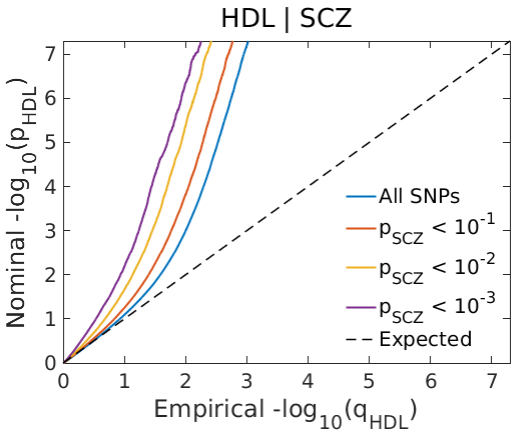
Figure S12. Polygenic overlap between SCZ and CVD phenotype. Conditional Q-Q plots of nominal versus empirical −log10p values (corrected for inflation) in SCZ below the standard GWAS threshold of $p < 5 \times 10{-}8$ as a function of significance of association with CVD phenotype, at the level of $p < 0.1$, $p < 0.01$, $p < 0.001$, respectively. The blue lines indicate all SNPs. The dashed lines indicate the null hypothesis. Abbreviations: SCZ, schizophrenia; CVD, cardiovascular disease; BMI, body mass index; TG, triglycerides; HDL, high-density lipoprotein cholesterol; LDL, low density lipoprotein cholesterol; TC, total cholesterol; SBP, systolic blood pressure; DBP, diastolic blood pressure; T2D, type 2 diabetes; CAD, coronary artery disease; WHR, waist-to-hip ratio adjusted for BMI. The conditional Q-Q plots build on the condFDR method.

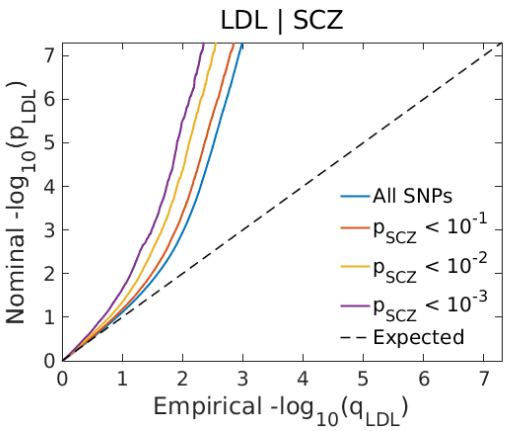

BMI | SCZ
Nominal -log10(pBMI)
Empirical -log10(qBMI)
All SNPs
pSCZ < 10-1
pSCZ < 10-2
pSCZ < 10-3
Expected

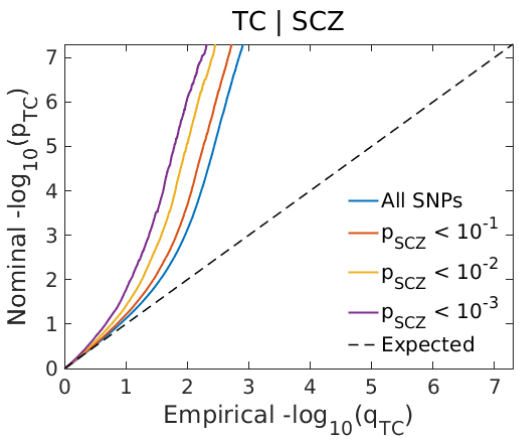

SmokingInitiation | SCZ
Nominal -log10(pSmokingInitiation)
Empirical -log10(qSmokingInitiation)
All SNPs
pSCZ < 10-1
pSCZ < 10-2
pSCZ < 10-3
Expected

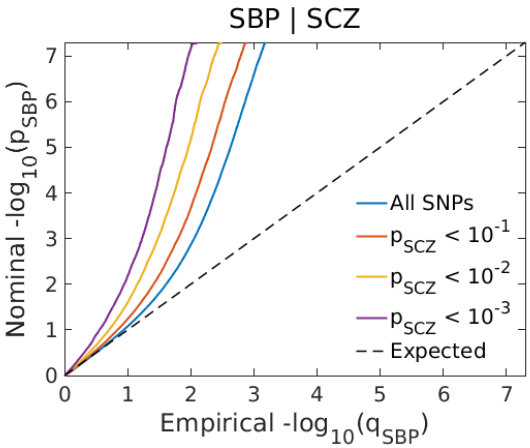

CigarettesPerDay | SCZ
Nominal -log10(pCigarettesPerDay)
Empirical -log10(qCigarettesPerDay)
All SNPs
pSCZ < 10-1
pSCZ < 10-2
pSCZ < 10-3
Expected


TG | SCZ
Nominal -log10(pTG)
Empirical -log10(qTG)
All SNPs
pSCZ < 10-1
pSCZ < 10-2
pSCZ < 10-3
Expected


HDL | SCZ
Nominal -log10(pHDL)
Empirical -log10(qHDL)
All SNPs
pSCZ < 10-1
pSCZ < 10-2
pSCZ < 10-3
Expected


LDL | SCZ
Nominal -log10(pLDL)
Empirical -log10(qLDL)
All SNPs
pSCZ < 10-1
pSCZ < 10-2
pSCZ < 10-3
Expected


TC | SCZ
Nominal -log10(pTC)
Empirical -log10(qTC)
All SNPs
pSCZ < 10-1
pSCZ < 10-2
pSCZ < 10-3
Expected


SBP | SCZ
Nominal -log10(pSBP)
Empirical -log10(qSBP)
All SNPs
pSCZ < 10-1
pSCZ < 10-2
pSCZ < 10-3
Expected

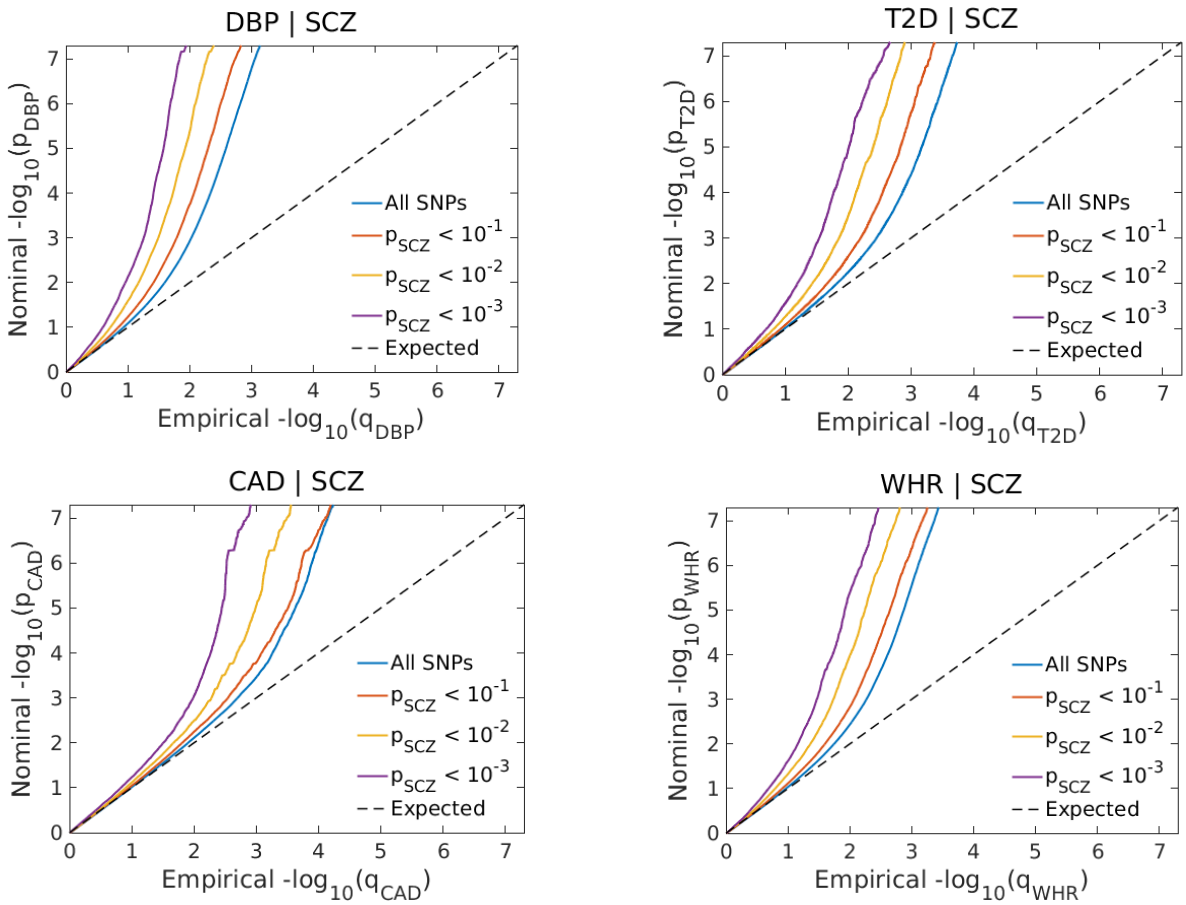


Figure S13. Polygenic overlap between SCZ and CVD phenotype. Conditional Q-Q plots of nominal versus empirical −log10p values (corrected for inflation) in CVD phenotype below the standard GWAS threshold of $p < 5 \times 10{-}8$ as a function of significance of association with SCZ, at the level of $p < 0.1$, $p < 0.01$, $p < 0.001$, respectively. The blue lines indicate all SNPs. The dashed lines indicate the null hypothesis. Abbreviations: SCZ, schizophrenia; CVD, cardiovascular disease; BMI, body mass index; TG, triglycerides; HDL, high-density lipoprotein cholesterol; LDL, low density lipoprotein cholesterol; TC, total cholesterol; SBP, systolic blood pressure; DBP, diastolic blood pressure; T2D, type 2 diabetes; CAD, coronary artery disease; WHR, waist-to-hip ratio adjusted for BMI. The conditional Q-Q plots build on the condFDR method.

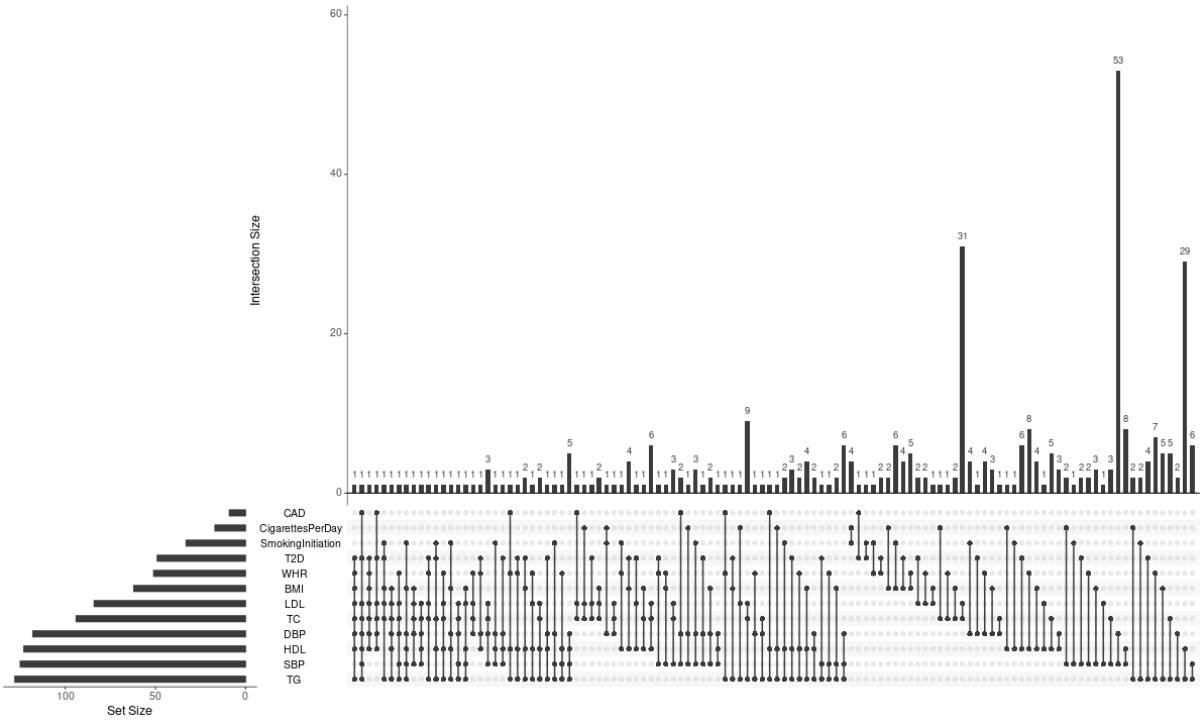


Figure S14. The figure provides an overview of the number of common lead SNPs for schizophrenia and more than one cardiovascular disease (CVD) phenotype at conjFDR<0.05. The bars to the right (vertical) provide information about the number of lead SNPs that are common to specific CVD phenotypes (e.g. BMI and lipids). The bars to the left (horizontal) show, for each CVD phenotypes, the number of lead SNPs that are found for at least one other CVD phenotype. Further information is available in Table S37. Abbreviations: CAD, coronary heart disease; T2D, type 2 diabetes; SBP, systolic blood pressure; DBP, diastolic blood pressure; BMI, body mass index; WHR, waist-to-hip ratio adjusted for BMI; HDL, high-density lipoprotein cholesterol; LDL, low- density lipoprotein cholesterol; TC, total cholesterol; TG, triglycerides.

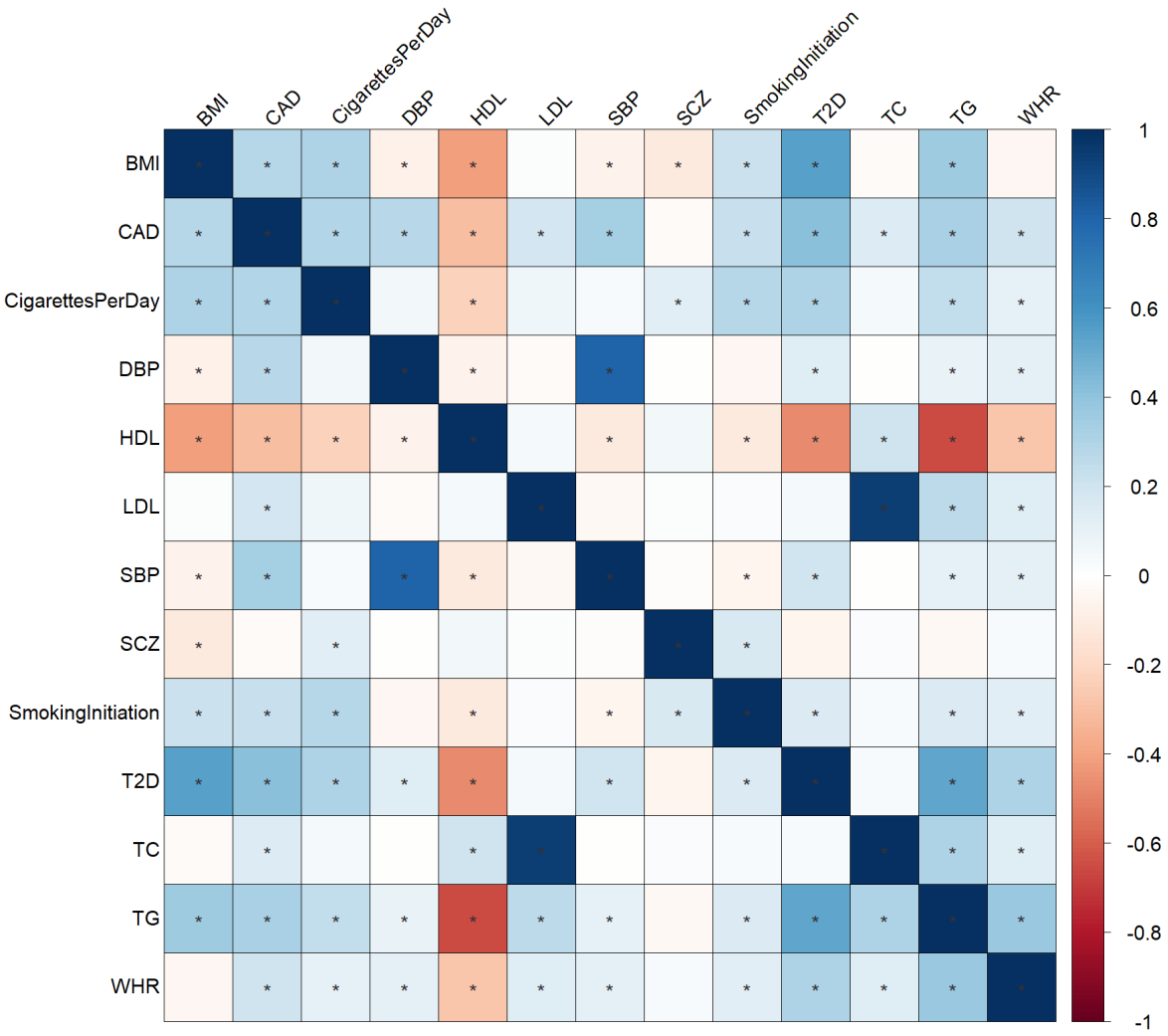
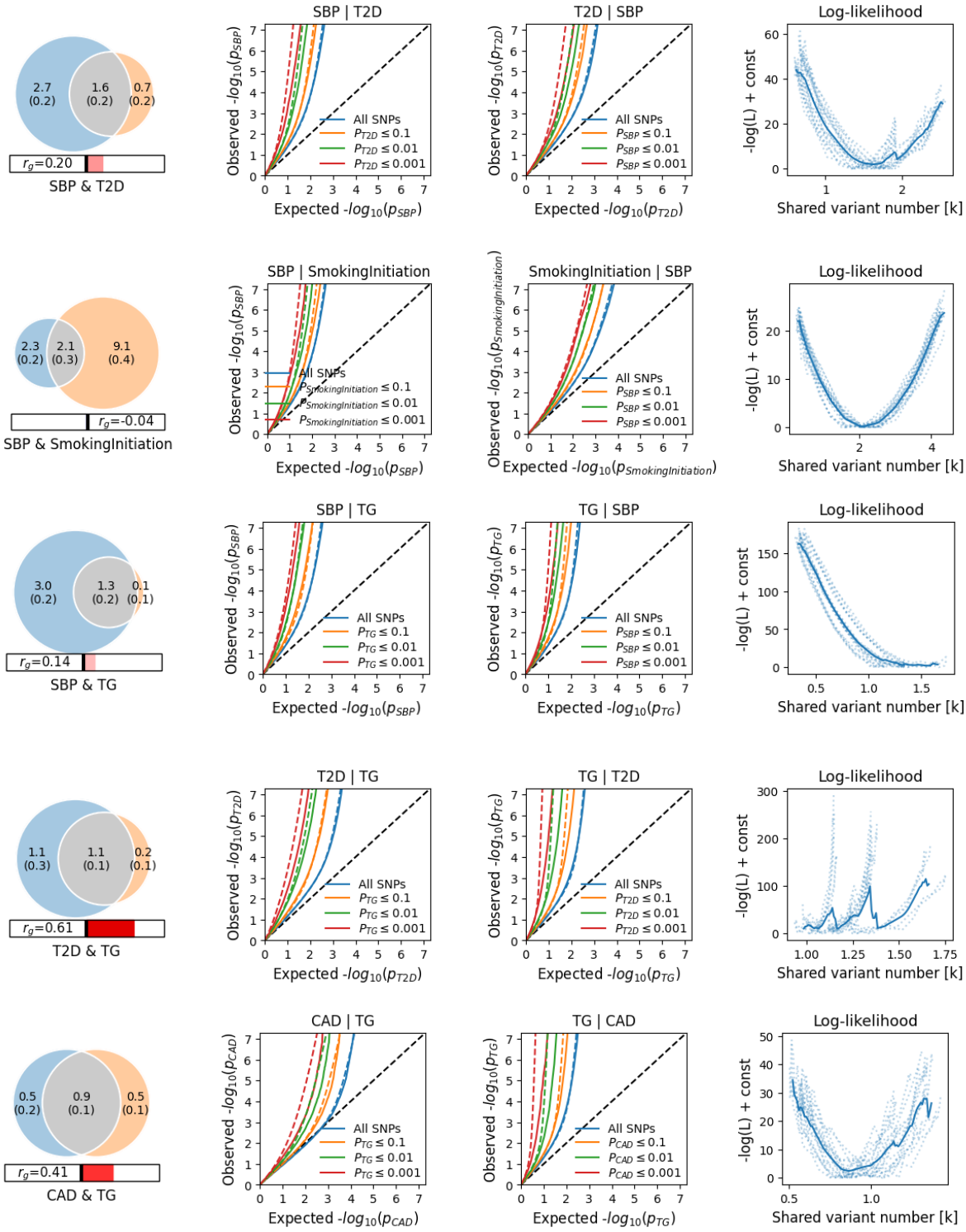
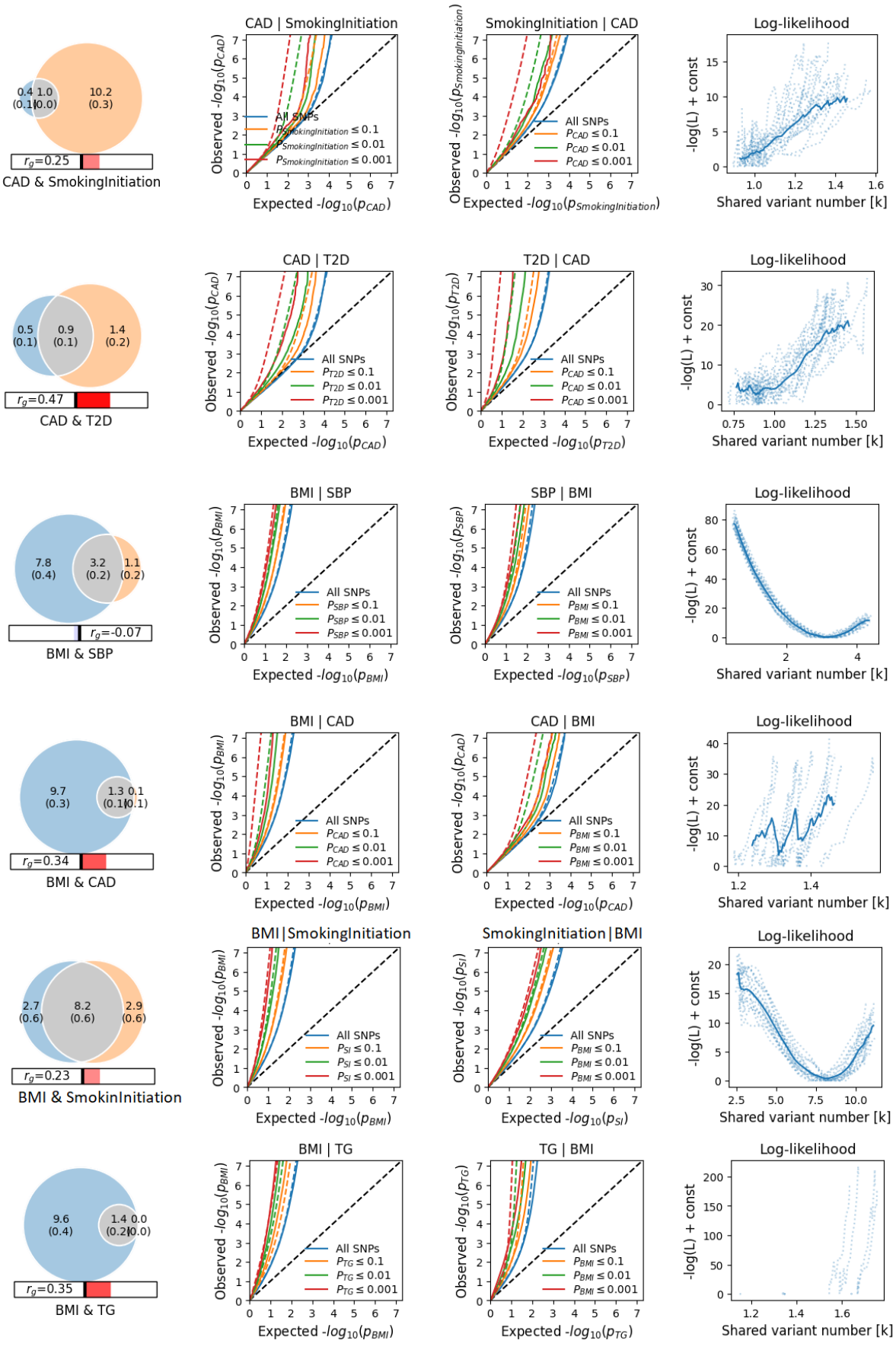


Figure S15. Matrix of the genetic correlations between CVD phenotypes, and between SCZ and CVD phenotypes. Blue denotes positive genetic correlation; red indicates negative genetic correlation. Genetic correlations that are significantly different from zero after Bonferroni correction are marked with an asterisk. Abbreviations: BMI, body mass index; CAD, coronary artery disease; DBP, diastolic blood pressure; SBP, systolic blood pressure; HDL, high-density lipoprotein cholesterol; LDL, low density lipoprotein cholesterol; TG, triglycerides; TC, total cholesterol; T2D, type 2 diabetes; WHR, waist-to-hip ratio adjusting for BMI; SCZ, schizophrenia; CVD, cardiovascular disease.


2.7 (0.2)
1.6 (0.2)
0.7 (0.2)
r_g=0.20
SBP & T2D
SBP | T2D
T2D | SBP
Log-likelihood
2.3 (0.2)
2.1 (0.3)
9.1 (0.4)
r_g=-0.04
SBP & SmokingInitiation
SBP | SmokingInitiation
SmokingInitiation | SBP
3.0 (0.2)
1.3 (0.2)
0.1 (0.1)
r_g=0.14
SBP & TG
SBP | TG
TG | SBP
1.1 (0.3)
1.1 (0.1)
0.2 (0.1)
r_g=0.61
T2D & TG
T2D | TG
TG | T2D
0.5 (0.2)
0.9 (0.1)
0.5 (0.1)
r_g=0.41
CAD & TG
CAD | TG
TG | CAD
All SNPs
Observed -log10(pSBP)
Expected -log10(pSBP)
Observed -log10(pT2D)
Expected -log10(pT2D)
Observed -log10(pSmokingInitiation)
Expected -log10(pSmokingInitiation)
Observed -log10(pTG)
Expected -log10(pTG)
Observed -log10(pCAD)
Expected -log10(pCAD)
-log(L) + const
Shared variant number [k]


0.4 (0.1)
1.0 (0.0)
10.2 (0.3)
r_g=0.25
CAD & SmokingInitiation
CAD | SmokingInitiation
SmokingInitiation | CAD
Log-likelihood
Observed -log10(pCAD)
Expected -log10(pCAD)
Observed -log10(pSmokingInitiation)
Expected -log10(pSmokingInitiation)
-log(L) + const
Shared variant number [k]
All SNPs
PSmokingInitiation ≤ 0.1
PSmokingInitiation ≤ 0.01
PSmokingInitiation ≤ 0.001
PCAD ≤ 0.1
PCAD ≤ 0.01
PCAD ≤ 0.001
0.5 (0.1)
0.9 (0.1)
1.4 (0.2)
r_g=0.47
CAD & T2D
CAD | T2D
T2D | CAD
Observed -log10(pT2D)
Expected -log10(pT2D)
PT2D ≤ 0.1
PT2D ≤ 0.01
PT2D ≤ 0.001
7.8 (0.4)
3.2 (0.2)
1.1 (0.2)
r_g=-0.07
BMI & SBP
BMI | SBP
SBP | BMI
Observed -log10(pBMI)
Expected -log10(pBMI)
Observed -log10(pSBP)
Expected -log10(pSBP)
PSBP ≤ 0.1
PSBP ≤ 0.01
PSBP ≤ 0.001
PBMI ≤ 0.1
PBMI ≤ 0.01
PBMI ≤ 0.001
9.7 (0.3)
1.3 (0.1)
0.1 (0.1)
r_g=0.34
BMI & CAD
BMI | CAD
CAD | BMI
Expected -log10(pCAD)
2.7 (0.6)
8.2 (0.6)
2.9 (0.6)
r_g=0.23
BMI & SmokinInitiation
BMI|SmokingInitiation
SmokingInitiation|BMI
PSI ≤ 0.1
PSI ≤ 0.01
PSI ≤ 0.001
9.6 (0.4)
1.4 (0.2)
0.0 (0.0)
r_g=0.35
BMI & TG
BMI | TG
TG | BMI
Observed -log10(pTG)
Expected -log10(pTG)
PTG ≤ 0.1
PTG ≤ 0.01
PTG ≤ 0.001

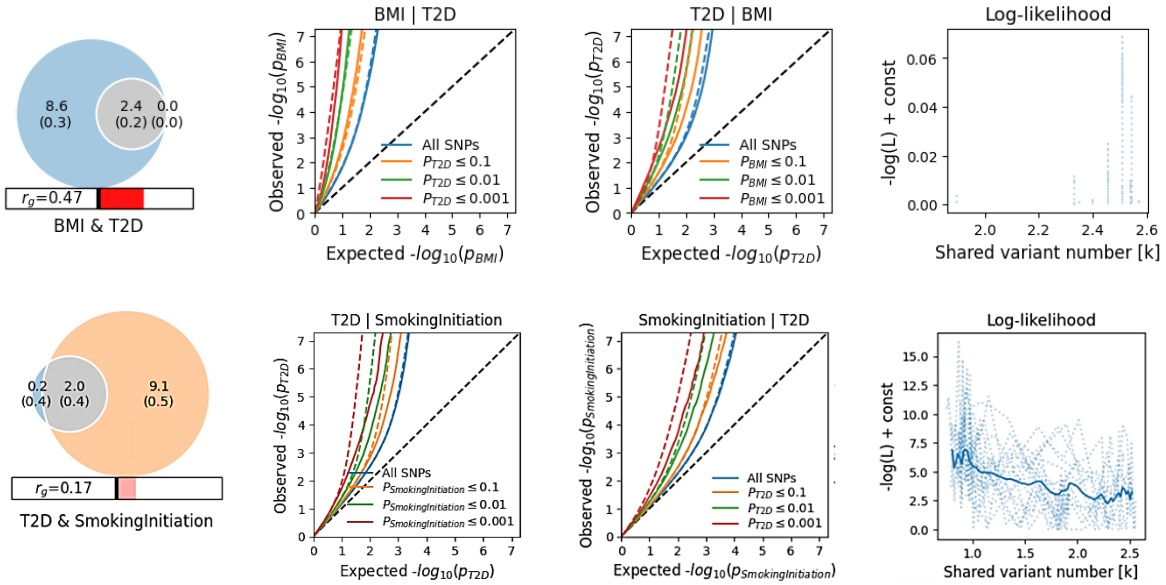
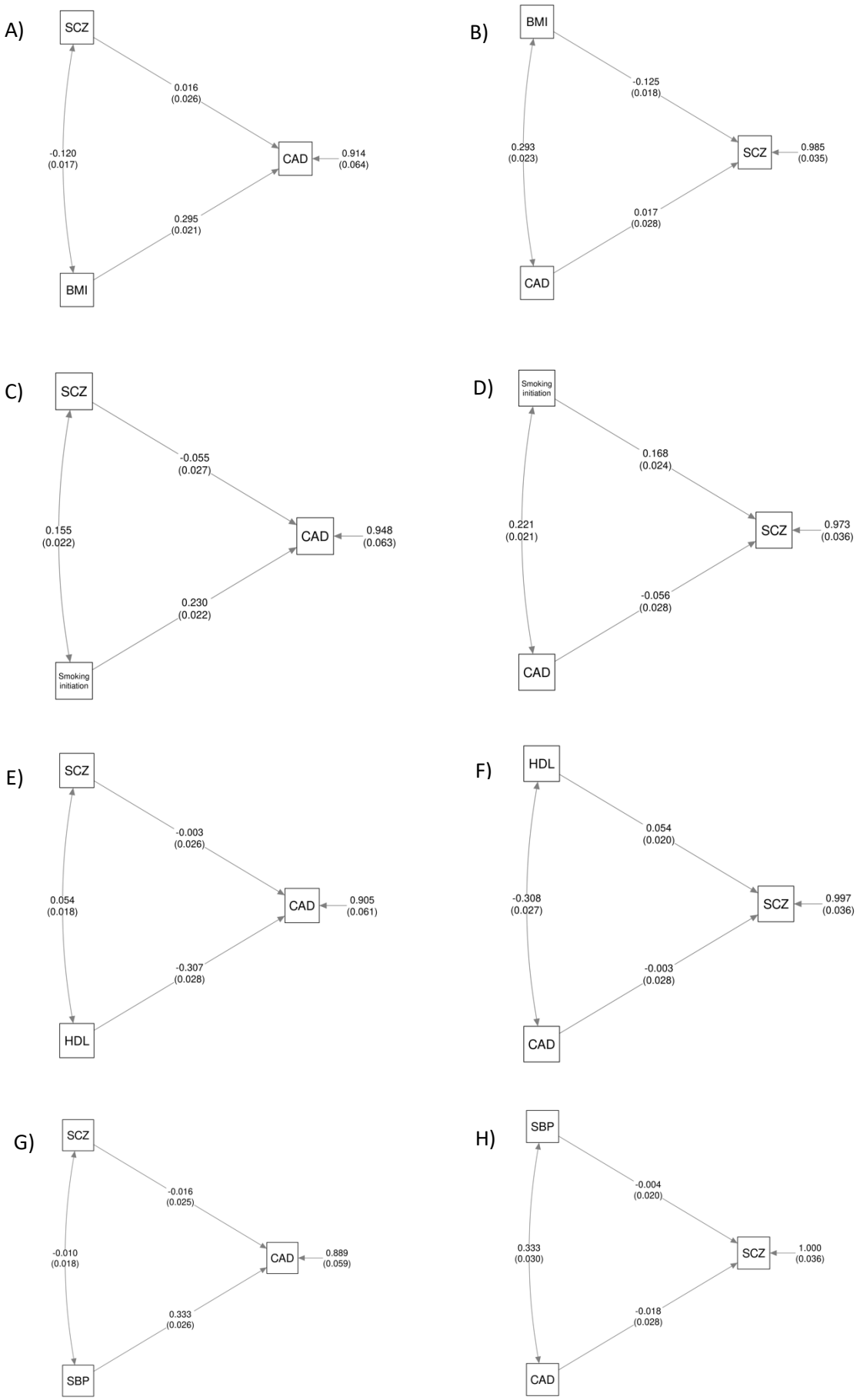


Figure S16. Venn Diagrams, conditional Q-Q plots, and negative log-likelihood plot, respectively. Venn diagrams of shared and unique trait-influencing variants, showing polygenic overlap (gray) between cardiovascular disease (CVD) phenotypes. The numbers in the Venn diagram indicate the estimated quantity of trait-influencing variants (in thousands), followed by standard error. Appearance of the Q-Q plot and negative log-likelihood plot are described below Figure S3. Figure generated from MiXeR. Abbreviations: BMI, body mass index; TG, triglycerides; SBP, systolic blood pressure; T2D, type 2 diabetes; CAD, coronary artery disease.


A)
SCZ
0.016
(0.026)
CAD
0.914
(0.064)
-0.120
(0.017)
0.295
(0.021)
BMI
B)
BMI
-0.125
(0.018)
SCZ
0.985
(0.035)
0.293
(0.023)
0.017
(0.028)
CAD
C)
SCZ
-0.055
(0.027)
CAD
0.948
(0.063)
0.155
(0.022)
0.230
(0.022)
Smoking initiation
D)
Smoking initiation
0.168
(0.024)
SCZ
0.973
(0.036)
0.221
(0.021)
-0.056
(0.028)
CAD
E)
SCZ
-0.003
(0.026)
CAD
0.905
(0.061)
0.054
(0.018)
-0.307
(0.028)
HDL
F)
HDL
0.054
(0.020)
SCZ
0.997
(0.036)
-0.308
(0.027)
-0.003
(0.028)
CAD
G)
SCZ
-0.016
(0.025)
CAD
0.889
(0.059)
-0.010
(0.018)
0.333
(0.026)
SBP
H)
SBP
-0.004
(0.020)
SCZ
1.000
(0.036)
0.333
(0.030)
-0.018
(0.028)
CAD

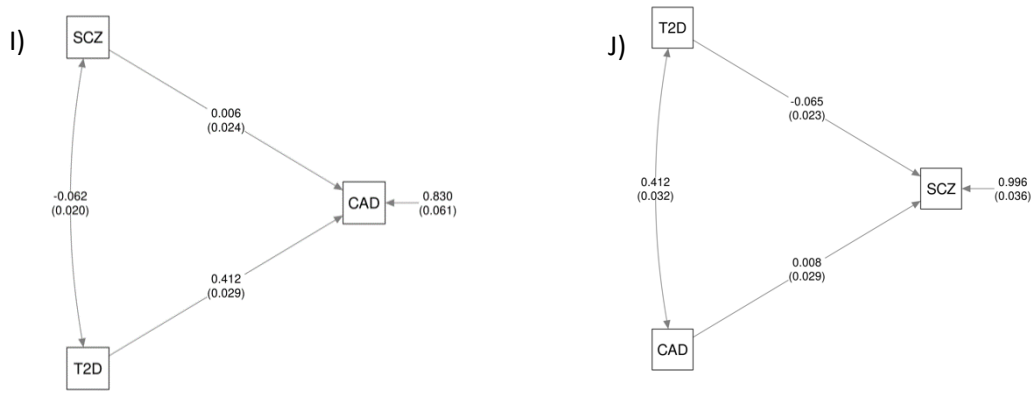


Figure S17. Genomic Structural Equation Modeling of schizophrenia (SCZ), coronary artery disease (CAD) and CVD risk factors. Paths diagrams for model A-I are presented: The path from SCZ to CAD (A) and vice versa (B) while controlling for body mass index (BMI). The path from SCZ to CAD (C) and vice versa (D) while controlling for smoking initiation. The path from SCZ to CAD (E) and vice versa (F) while controlling for high-density lipoprotein (HDL). The path from SCZ to CAD (G) and vice versa (H) while controlling for systolic blood pressure (SBP). The path from SCZ to CAD (I) and vice versa (J) while controlling for type 2 diabetes (T2D). Single-headed arrows represent a partial correlation, while double-headed arrows represent a bivariate correlation calculated with LDSR. The residual variance is shown to the right in each diagram.

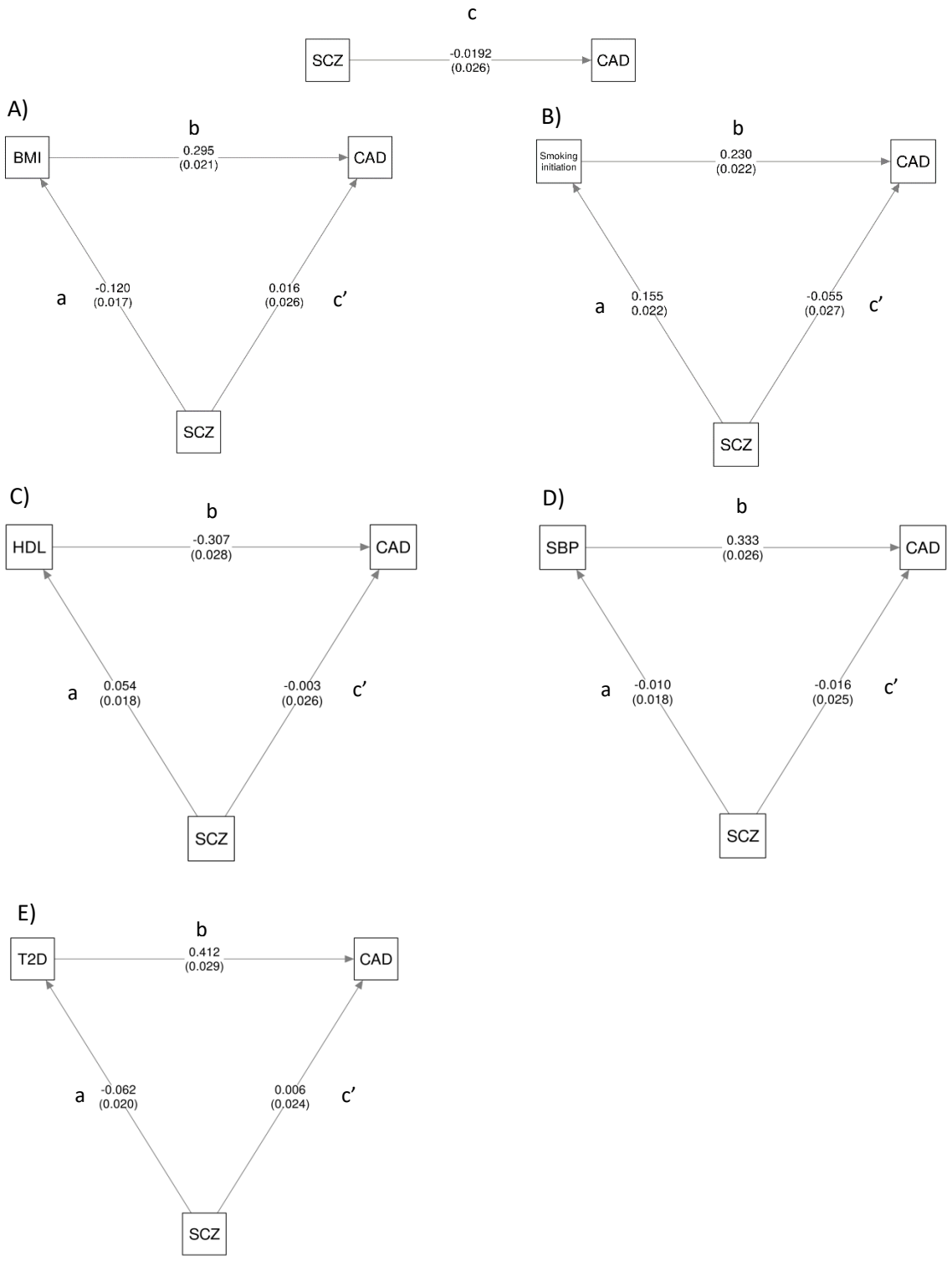


Figure S18. Mediation analyses with Genomic Structural Equation Modeling. Path diagrams with CVD risk factors as mediators of the relationship between schizophrenia (SCZ) and coronary artery disease (CAD), including factor loadings for each path. Standard errors are included in parenthesis. Each path is labeled, where c is the total effect for the path between SCZ and CAD, and c' is the direct path from SCZ to CAD, and the indirect effect is the product of paths a and b (a*b). Models assessing

the effect of body mass index (BMI) (A), smoking initiation (B), high-density lipoprotein (HDL) (C), systolic blood pressure (SBP) (D) and type 2 diabetes (T2D) (E).

References

1. Trubetskoy V, Pardiñas AF, Qi T, et al. Mapping genomic loci implicates genes and synaptic biology in schizophrenia. Nature. 2022.
2. Turcot V, Lu Y, Highland HM, et al. Protein-altering variants associated with body mass index implicate pathways that control energy intake and expenditure in obesity. Nat Genet. 2018;50:26-41.
3. Pulit SL, Stoneman C, Morris AP, et al. Meta-analysis of genome-wide association studies for body fat distribution in 694 649 individuals of European ancestry. Hum Mol Genet. 2019;28:166-174.
4. Graham SE, Clarke SL, Wu K-HH, et al. The power of genetic diversity in genome-wide association studies of lipids. Nature. 2021.
5. Evangelou E, Warren HR, Mosen-Ansorena D, et al. Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. Nature genetics. 2018;50:1412-1425.
6. Mahajan A, Taliun D, Thurner M, et al. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. Nature Genetics. 2018;50:1505-1513.
7. Liu M, Jiang Y, Wedow R, et al. Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. Nature Genetics. 2019;51:237-244.
8. Nelson CP, Goel A, Butterworth AS, et al. Association analyses based on false discovery rate implicate new loci for coronary artery disease. Nat Genet. 2017;49:1385-1391.
9. Frei O, Holland D, Smeland OB, et al. Bivariate causal mixture model quantifies polygenic overlap between complex traits beyond genetic correlation. Nature communications. 2019;10:2417-2417.
10. Holland D, Frei O, Desikan R, et al. Beyond SNP heritability: Polygenicity and discoverability of phenotypes estimated with a univariate Gaussian mixture model. PLOS Genetics. 2020;16:e1008612.
11. Bulik-Sullivan BK, Loh P-R, Finucane HK, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nature Genetics. 2015;47:291-295.
12. Benjamini Y, Hochberg Y: Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. in Journal of the Royal Statistical Society Series B (Methodological), Blackwell Publishing; 1995. pp. 289-300.
13. Efron B. Size, power and false discovery rates. The Annals of Statistics. 2007;35:1351–1377.
14. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. American journal of human genetics. 2007;81:559-575.
15. Schweder T, Spjotvoll E. Plots of P-Values to Evaluate Many Tests Simultaneously. Biometrika. 1982;69:493-502.
16. Andreassen OA, Djurovic S, Thompson WK, et al. Improved detection of common variants associated with schizophrenia by leveraging pleiotropy with cardiovascular-disease risk factors. Am J Hum Genet. 2013;92:197-209.
17. Andreassen OA, Thompson WK, Schork AJ, et al. Improved detection of common variants associated with schizophrenia and bipolar disorder using pleiotropy-informed conditional false discovery rate. PLoS genetics. 2013;9:e1003455.
18. Andreassen OA, Harbo HF, Wang Y, et al. Genetic pleiotropy between multiple sclerosis and schizophrenia but not bipolar disorder: differential involvement of immune-related gene loci. Molecular psychiatry. 2015;20:207-214.
19. Andreassen OA, Thompson WK, Dale AM. Boosting the power of schizophrenia genetics by leveraging new statistical tools. Schizophrenia bulletin. 2014;40:13-17.

20. Andreassen OA, Desikan RS, Wang Y, et al. Abundant genetic overlap between blood lipids and immune-mediated diseases indicates shared molecular genetic mechanisms. PloS one. 2015;10:e0123057.
21. Nichols T, Brett M, Andersson J, et al. Valid conjunction inference with the minimum statistic. Neuroimage. 2005;25:653-660.
22. Schwartzman A, Lin X. The effect of correlation in false discovery rate estimation. Biometrika. 2011;98:199-214.
23. Smeland OB, Frei O, Shadrin A, et al. Discovery of shared genomic loci using the conditional false discovery rate approach. Human Genetics. 2020;139:85-94.
24. Rodevand L, Bahrami S, Frei O, et al. Extensive bidirectional genetic overlap between bipolar disorder and cardiovascular disease phenotypes. Transl Psychiatry. 2021;11:407.
25. Bahrami S, Steen NE, Shadrin A, et al. Shared Genetic Loci Between Body Mass Index and Major Psychiatric Disorders: A Genome-wide Association Study. JAMA Psychiatry. 2020;77:503-512.
26. Frei O, Holland D, Smeland OB, et al. Bivariate causal mixture model quantifies polygenic overlap between complex traits beyond genetic correlation. Nature Communications. 2019;10:2417.
27. Watanabe K, Taskesen E, van Bochoven A, et al. Functional mapping and annotation of genetic associations with FUMA. Nat Commun. 2017;8:1826.
28. The 1000 Genomes Project Consortium. A global reference for human genetic variation. Nature. 2015;526:68-74.
29. Bulik-Sullivan B, Finucane HK, Anttila V, et al. An atlas of genetic correlations across human diseases and traits. Nature Genetics. 2015;47:1236-1241.
30. Buniello A, MacArthur JA L, Cerezo M, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Research. 2018;47:D1005-D1012.
31. Bahrami S, Hindley G, Winsvold BS, et al. Dissecting the shared genetic basis of migraine and mental disorders using novel statistical tools. Brain. 2021;145:142-153.
32. Wiström ED, O'Connell KS, Karadag N, et al. Genome-wide analysis reveals genetic overlap between alcohol use behaviours, schizophrenia and bipolar disorder and identifies novel shared risk loci. Addiction. 2022;117:600-610.
33. Kircher M, Witten DM, Jain P, et al. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet. 2014;46:310-315.
34. Boyle AP, Hong EL, Hariharan M, et al. Annotation of functional variation in personal genomes using RegulomeDB. Genome Res. 2012;22:1790-1797.
35. Roadmap Epigenomics C, Kundaje A, Meuleman W, et al. Integrative analysis of 111 reference human epigenomes. Nature. 2015;518:317-330.
36. Zhu Z, Zhang F, Hu H, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. Nat Genet. 2016;48:481-487.
37. Schmitt AD, Hu M, Jung I, et al. A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome. Cell Rep. 2016;17:2042-2059.
38. Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000;25:25-29.
39. Liberzon A, Subramanian A, Pinchback R, et al. Molecular signatures database (MSigDB) 3.0. Bioinformatics. 2011;27:1739-1740.
40. Uffelmann E, Huang QQ, Munung NS, et al. Genome-wide association studies. Nature Reviews Methods Primers. 2021;1:59.
41. Wormington B, Thorp JG, Scott JG, et al. Influences on the Genetic Relationship Between Cannabis Use and Schizophrenia: The Role of the Externalizing Spectrum. Schizophr Bull. 2022;48:1318-1326.
42. Grotzinger AD, Rhemtulla M, de Vlaming R, et al. Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits. Nat Hum Behav. 2019;3:513-525.

43. Schoech AP, Jordan DM, Loh P-R, et al. Quantification of frequency-dependent genetic architectures in 25 UK Biobank traits reveals action of negative selection. Nature Communications. 2019;10:790.
44. Speed D, Balding DJ. SumHer better estimates the SNP heritability of complex traits from summary statistics. Nature Genetics. 2019;51:277-284.
45. Zeng J, de Vlaming R, Wu Y, et al. Signatures of negative selection in the genetic architecture of human complex traits. Nature Genetics. 2018;50:746-753.
46. Zeng J, Xue A, Jiang L, et al. Widespread signatures of natural selection across human complex traits and functional genomic categories. Nature Communications. 2021;12:1164.
47. Kenny DA, Judd CM. Power Anomalies in Testing Mediation. Psychological Science. 2013;25:334-339.